
Theses and Dissertations

Fall 2011

Second language reading topic familiarity and test score: test-taking strategies for multiple-choice comprehension questions

Jia-Ying Lee
University of Iowa

Copyright 2011 Jia-Ying Lee

This dissertation is available at Iowa Research Online: <http://ir.uiowa.edu/etd/2737>

Recommended Citation

Lee, Jia-Ying. "Second language reading topic familiarity and test score: test-taking strategies for multiple-choice comprehension questions." PhD (Doctor of Philosophy) thesis, University of Iowa, 2011.
<http://ir.uiowa.edu/etd/2737>.

Follow this and additional works at: <http://ir.uiowa.edu/etd>



Part of the [Teacher Education and Professional Development Commons](#)

**SECOND LANGUAGE READING TOPIC FAMILIARITY AND TEST SCORE:
TEST-TAKING STRATEGIES FOR MULTIPLE-CHOICE
COMPREHENSION QUESTIONS**

by

Jia-Ying Lee

An Abstract

Of a thesis submitted in partial fulfillment of the requirements
for the Doctor of Philosophy degree in
Teaching and Learning (Foreign Language and ESL Education)
in the Graduate College of The University of Iowa

December 2011

Thesis Supervisor: Assistant Professor Lia Plakans

ABSTRACT

The main purpose of this study was to compare the strategies used by Chinese-speaking students when confronted with familiar versus unfamiliar topics in a multiple-choice format reading comprehension test. The focus was on describing what students do when they are taking reading comprehension tests by asking students to verbalize their thoughts. The strategies were further compared with participants' level of familiarity with different reading topics and their reading scores.

Twenty Chinese-speaking participants at the University of Iowa performed three tasks: a topical knowledge vocabulary assessment that served as an indicator of each participant's topical knowledge about the four selected content areas in this study (law, business, language teaching, and engineering); two Test of English as a Foreign Language (TOEFL) internet-based test (iBT) practice reading comprehension passages, one with a familiar topic and the other with an unfamiliar topic, and both with retrospective think-aloud protocols; and an interview related to participants' test-taking strategies.

Two stages of analysis, qualitative and quantitative, were undertaken in this study. For the qualitative analysis, all verbal reports provided by participants in the think-aloud protocols and the interviews were recorded and transcribed. Six categories of strategies emerged: general approaches to reading the passages, identification of important information by the discourse structure of the passages, vocabulary/sentence-in-context approaches, multiple-choice test-management strategies, test-wiseness, and background knowledge.

For the quantitative analysis, an analysis of variance (ANOVA) with repeated measures was completed to determine if there were significant differences based on the frequency of strategy use and level of topic familiarity. The results showed that the types of test-taking strategies adopted by Chinese-speaking graduate students remained similar when they read passages with familiar versus unfamiliar topics. However, participants all

reported feeling more relief and more confidence when reading passages were related to their background knowledge.

The second ANOVA employed a split-plot statistical design to examine whether there were significant differences based on participants' strategy use and their reading scores as measured by the iBT reading comprehension tests. High scorers employed strategies in categories one, two, three, and four significantly more frequently than low scorers. However, low scorers adopted significantly more strategies in category five than high scorers. In category six, high and low scorers seemed to use a similar number of strategies.

Findings that emerged from the two perspectives are discussed; implications related to test-taking and reading pedagogy are provided in the conclusion.

Abstract Approved: _____

Thesis Supervisor

Title and Department

Date

**SECOND LANGUAGE READING TOPIC FAMILIARITY AND TEST SCORE:
TEST-TAKING STRATEGIES FOR MULTIPLE-CHOICE
COMPREHENSION QUESTIONS**

by

Jia-Ying Lee

A thesis submitted in partial fulfillment of the requirements
for the Doctor of Philosophy degree in
Teaching and Learning (Foreign Language and ESL Education)
in the Graduate College of The University of Iowa

December 2011

Thesis Supervisor: Assistant Professor Lia Plakans

Graduate College
The University of Iowa
Iowa City, Iowa

CERTIFICATE OF APPROVAL

PH.D. THESIS

This is to certify that the Ph.D. thesis of

Jia-Ying Lee

has been approved by the Examining Committee for the thesis requirement for the Doctor of Philosophy degree in Teaching and Learning (Foreign Language and ESL Education) at the December 2011 graduation.

Thesis Committee:

Lia Plakans, Thesis Supervisor

Michael Everson

Leslie Schrier

Pamela Wesely

Timothy Ansley

To my father, Mr. Tche-ming Lee, 李哲明先生

To my mother, Mrs. Su-o Sun, 孫素娥女士

ACKNOWLEDGMENTS

This dissertation would not have been completed without the help of many people. First of all, I would like to express my sincerest gratitude and appreciation to my dissertation supervisor, Dr. Lia Plakans, for her inspiration and her invaluable assistance during my dissertation writing journey. I am extremely fortunate and grateful to have her constant guidance and continual support. Being both a great class instructor and research advisor has inspired me to become an influential teacher in my future career. Without her direction and encouragement, I would not have been able to complete my graduate study.

I would also like to thank all of my committee members: Dr. Michael Everson, who mentored me since the day I started my graduate study at the University of Iowa; Dr. Pamela Wesely, for her valuable and generous guidance in the research methodology; Dr. Timothy Ansley, for his feedback related to my statistical analysis; Dr. Leslie Schrier, for her insights to my study. I greatly appreciate the time and effort they contributed to serving on my dissertation committee. A special note of thanks goes to Dr. Sheila Barron and Yi-fang Wu for their statistical support and assistance in the Statistics Outreach Center at the UI.

Additionally, my appreciation goes to all the individuals who responded to my study invitation and those who participated in the study. Without their generous participation, the completion of this dissertation would not have been possible.

The Small Grant for Doctoral Research in Second or Foreign Language Assessment awarded from the Educational Testing Service and the Summer Graduate Fellowship from the University of Iowa provided financial support to this study, as well as gave me confidence in my research.

I also express my utmost gratitude to these fellow graduate students in my program: Ho Jung Choi, Cindra Visser, Fatima Baig, Aurore Mroz, and Tingting Chen. I am extremely lucky to have Ho Jung and Cindra as my wonderful

companions while pursuing my doctorate. The laughter and tears we shared through these thousands of days are unforgettable. The numerous help and warm pep talks received from Fatima and Aurore greatly reduced my anxiety. And thanks to Tingting, who helped me with data analysis by offering her valuable perspectives and generous time.

I would also like to offer thanks to my Taiwanese friends who provided warm friendship and encouragement: Hsiu-ying Chiang, Chu-Ling Lo, Shanju Lin, Sih-han Janice Wang, and Yi-fang Wu. Special thanks to Kuan-hua Chen, who gave me emotional support, and his generous love made this dissertation process much more enjoyable than if I had to do it alone.

Last but not least, I am indebted to my parents, for their unconditional and wholehearted love and support, both spiritually and financially. Their encouragement through Skype for these past five years and their belief in me were the motivation for this dissertation. My gratitude to them is beyond words.

TABLE OF CONTENTS

LIST OF TABLES	vii
LIST OF FIGURES	viii
CHAPTER	
1. INTRODUCTION.....	1
Topic Familiarity in Reading Comprehension	2
Reading Strategies.....	2
Strategies for Readers with Different Cultural Backgrounds.....	3
TOEFL Internet-based Test.....	3
Multiple-choice Items.....	5
Problem Statement	6
Purpose of the Study	6
Significance of the Study	7
Primary Research Questions	7
Definition of Terms.....	8
2. LITERATURE REVIEW	9
The Nature of Reading	9
The Important Role of the Reader	10
Topic Familiarity in Reading Comprehension	12
The Impact of Content Knowledge	14
The Impact of Formal Knowledge	18
Strategies in Second Language Reading Comprehension.....	20
Strategies with Different L2 Language Proficiency.....	23
Strategies with Readers' L1 Pedagogical Cultural Background	26
Second Language Reading Comprehension Assessment.....	27
Validity and Reading Comprehension Tests	28
Multiple-choice Format in Reading Comprehension Assessment	30
Multiple-choice Format and Topic Familiarity	32
Multiple-choice Items and Strategy Use	34
Conclusion.....	38
3. METHODOLOGY.....	40
Research Design.....	40
Research Questions	42
Participants	42
Materials.....	43
Data Collection Instruments.....	45
Topical Knowledge Vocabulary Assessment.....	45
Think-Aloud Protocols	49
Interviews	51
Data Collection Procedure	51
Informational Session: Introduction, Retrospective Think-Aloud Practice, and Topical Knowledge Vocabulary Assessment	52
Test-Taking Session: Practice Test and Retrospective Thinking-Aloud.....	53
Topical Knowledge and Reading Comprehension Assessment Results	54
Data Analysis	56

Qualitative Data Analysis.....	59
Quantitative Data Analysis.....	60
Pilot Study.....	61
4. RESULTS	63
Research Questions	63
Strategies and Topic Familiarity	63
Distribution of Strategy Use.....	64
Relationship Between Strategy Use and Topic Familiarity	72
Strategies and Test Score	81
Relationship Between Strategy Use and Test Score.....	81
Case Studies of High and Low Scorers	85
Participants' Perception of Test Questions.....	94
Summary	95
5. DISCUSSION OF FINDINGS AND CONCLUSION.....	96
Summary of this Study.....	96
Strategies Compared with Topic Familiarity	97
Strategies Compared with Test Scores.....	97
Discussion	98
Similar Strategies Across Topic Familiarity	98
Different Strategies Across Test Score.....	100
Differences in Metacognitive Awareness.....	100
Differences in Formal Knowledge Applications.....	102
The Influence of Chinese Pedagogical Cultural Background	102
The Impact of L2 Vocabulary on Test-taking Strategies	103
Multiple-choice Test Management Strategies	107
Test-wiseness Strategies.....	108
Implications for Test-taking and Pedagogy	110
Limitations and Suggestions for Future Study.....	112
Conclusion.....	113
REFERENCES	114
APPENDIX A. STUDY INTRODUCTORY FORM	122
APPENDIX B. PERMISSION LETTER FROM HARVARD PRESS	123
APPENDIX C. PERMISSION LETTER FROM JINNI PUBLISHING CORPORATION.....	124
APPENDIX D. iBT PRACTICE TEST – CIVIL LAW	125
APPENDIX E. iBT PRACTICE TEST – LAW OF DEMAND	129
APPENDIX F. TOPICAL KNOWLEDGE VOCABULARY ASSESSMENT.....	134
APPENDIX G. INTERVIEW QUESTIONS	135
APPENDIX H. VOLUNTARY SIGN-UP SHEET.....	136

LIST OF TABLES

Table

1. Participant Information	43
2. TOEFL iBT Questions and Examples	45
3. Information About Content Advisors	47
4. Results of Topical Knowledge Vocabulary Assessment	57
5. Results of Combined iBT Reading Comprehension Tests	58
6. TOEFL iBT Six Broad Categories of Strategies.....	73
7. Means and Standard Deviations of Familiar and Unfamiliar Categories of Strategies	75
8. ANOVA Analysis of Topic Familiarity and Reading Strategy	77
9. Results of the Tukey Follow-up Analysis.....	79
10. Means and Standard Deviations of Strategies by High and Low Scorers	82
11. Results of the 2 x 6 Split-plot ANOVA for Reading Scores and Strategy Use	82
12. Results of Follow-up Simple Effects Tests.....	83

LIST OF FIGURES

Figure

1. Bernhardt's (1986) constructivist model	12
2. Data collection procedures.....	55
3. Six categories of strategies used in familiar versus unfamiliar texts.....	76
4. Six categories of strategies used by high and low scorers.....	84

CHAPTER 1

INTRODUCTION

The ability to read and to decode written words effectively is vital to functioning in everyday life. In school, reading ability is viewed as critical to academic success because students read to learn and acquire new information (Alsheikh, 2011; Grabe & Stoller, 2002). Therefore, researchers have argued that the most important and essential skill in second or foreign language learning is reading (Bernhardt, 2005; Upton & Lee-Thompson, 2001). Given this importance, reading ability is assessed in many high-stakes language exams, such as the Test of English as a Foreign Language (TOEFL), the Test of English as International Communication (TOEIC), and the International English Language Testing System (IELTS).

The active role of readers cannot be overemphasized when it comes to reading comprehension, which is now viewed as the interaction between the reader and the writer. Accordingly, reading comprehension includes both the information in the printed text and the interpretation by the readers when they relate new textual information to information already stored in their memories (Bernhardt, 1986; Grabe, 2009; Kim, 2010; Radojevic, 2006). As Spiro (1980) concluded, “although text constrains the possible meaning, readers of different knowledge, interests and perspectives, or the same reader in different contexts, may construct quite different interpretations” (p. 32). Therefore, the meaning of a text does not only reside in the displayed text. Readers, instead, contribute more information than the printed words by using their prior knowledge with different life experiences. It has been observed that successful readers are strategic because they use the text with their prior knowledge as clues to construct meaning.

This study centered on readers and their strategies in Second Language (L2) reading when they completed simulated high-stakes multiple-choice test questions. Their strategies were further compared with (a) their familiarity with different topics and (b) their reading scores.

Topic Familiarity in Reading Comprehension

In recognizing the readers in the reading process, researchers in both first language (L1) and second language (L2) acquisition have investigated the important role of readers' prior knowledge in reading outcomes (Carrell, 1987; Chen & Donin, 1997; Krekeler, 2006; Lesser, 2003; Shapiro, 2004; Tsui, 2002; Wu, 2005). Students reading about topics with which they are familiar comprehend and recall more important and correct textual information better than those who are unfamiliar with the content of the text, which indicates that prior knowledge exerts a positive effect on measures of reading comprehension. It has been argued that with the assistance of prior knowledge, "readers at a lower level of [second] language proficiency could perform better than, or at least as well as, readers at a higher level of language proficiency" (Tsui, 2002, p. 29). In addition, prior knowledge helps readers to fill in the gaps when information in the text is not explicitly stated (Alderson, 2000; Leiser, 2003; Lin, 2002). In this way, readers draw inferences using prior knowledge related to the content to decode ambiguous messages in a text. Therefore, topic familiarity helps readers to contextualize textual ideas and facilitates their comprehension.

Reading Strategies

Reading strategies, which are conceptualized as "intentional actions chosen to facilitate reading at any level of processing" (Erler & Finkbeiner, 2007, p. 189), are central to comprehension and have been a focus of research for the past two decades (Alderson, 2000; Block, 1986, 1992; Cohen & Upton, 2007; Pritchard, 1990; Purpura, 1997; Radojevic, 2006). L2 reading researchers have determined that language proficiency accounts for differences in strategy use (Anderson, 1991; Block, 1986; Campbell, 1999; Logie, 1995; Ozuru et al., 2007). Therefore, a reader's ability to adopt the appropriate strategies and the ability to use these strategies proficiently have been recognized as important for increasing reading comprehension. To understand what skilled readers do when they read, strategy use across different proficiency levels has

been investigated (Alsheikh, 2011; Bang & Zhao, 2007; Block, 1992; Phakiti, 2003; Purpura, 1999). For example, proficient readers are reported to have more awareness of how to approach reading tasks and to monitor their comprehension. Conversely, less proficient readers have been found to use inappropriate reading strategies during reading tasks. Based on the importance of reading strategies, researchers have suggested that a top priority for reading instructors should be to teach their students when and how to use strategies more effectively to maximize reading comprehension (Block, 1992; Grabe, 2009; Wu, 2005).

Strategies for Readers with Different Cultural Backgrounds

Reading in a foreign/second language is a complex internal process, and the process can vary due to different individual factors; one of them is the readers' cultural background. Readers' cultural backgrounds influence reading strategy selection because readers have learned to read within the context of a particular culture with different educational experiences (Abbott, 2006; Block, 1986; Eler & Finkbeiner, 2007; Parry, 1996). Thus, English as second /foreign language learners who are from different cultural backgrounds with different L1s may use dramatically different reading strategies when comprehending the same English text. This statement supports the contention that the approach and comprehension of L2 reading depends heavily on readers' L1 sociocultural backgrounds because "L1 serves as a tool to help students think about and make sense...of L2 texts" (Upton & Lee-Thompson, 2001, p. 491). Thus, readers' cultural backgrounds should not be overlooked when exploring the reading strategies that different readers adopt.

TOEFL Internet-based Test

For international students who apply to schools in North America, a score on the Test of English as a Foreign Language (TOEFL) is one of the requirements in the admissions decision-making process (Sawaki, Stricker, & Oranje, 2009, p. 5). Since the Chinese believe that one becomes a more competitive candidate in his/her career if he/she

pursues an American degree, studying in the U.S. has become increasingly popular with Mandarin Chinese-speaking students. Also, with the rapid economic growth in China and Taiwan, most families can afford the financial expenses of studying in the U. S. The number of Chinese students who have taken the TOEFL exam and applied to schools in the U.S. has tripled recently. Thus, the Chinese-speaking students are an important population to study, who now represent the largest international student population at the University of Iowa in 2010 (1312 from China and 109 from Taiwan for a total of 2982). This population affects the demographics of universities in the U.S. Of all the international students who enrolled in American universities in 2009-2010 academic year, nearly 18% were Chinese-speaking students, according to the Institute of International Educations' Open Doors 2010 report.

The Internet-based TOEFL (iBT) Test, the newest version of the TOEFL, was introduced globally in late 2005. The TOEFL iBT is very different from the previous TOEFL Computer-based Test (CBT) due to its design, which intends to “reflect current theories of communicative language use in an academic context” (Taylor & Angelis, 2008, p. 41). With the addition of a speaking section and deletion of the grammar structure section, as well as changes in the listening and reading sections, the TOEFL iBT is intended to be more comparable to the various language tasks that test-takers encounter in their academic careers.

Although the TOEFL iBT reading sections still measure test-takers' reading ability using multiple-choice questions, the iBT has fewer but longer texts than the previous versions (600-700 vs. 300-400; Alderson, 2009). Each iBT reading test includes from three to five passages, and test-takers are allowed between 60 and 100 minutes to complete the reading section. The reading passages are adopted from university-level textbooks on different topics and consist of three types of text: exposition, argumentation, and historical narrative. Each passage is followed by 12 to 14 multiple-choice questions, which are designed to measure basic comprehension, inferencing, and reading to learn.

The reading to learn items are followed by multiple-choice questions with multiple correct answers for which test-takers receive partial credit if they do not answer all of them correctly.

Multiple-choice Items

Multiple-choice items are a commonly implemented format of reading assessment because of their practicality and scoring efficiency, with a quantifiable number that allows easy comparison across test-takers, especially in large-scale standardized exams. In addition, multiple-choice responses are machine-scoreable, which “minimizes the disadvantages inherent in assessment procedure that required subjective rating” (Campbell, 1999). However, several criticisms of multiple-choice tests include over-dependence on statistical values and test-wiseness strategies, which engage test-takers in a guessing game. An additional objection to multiple-choice tests brought up by some opponents is that test-takers select one correct answer from the options. This may limit test-takers’ opportunities to apply background knowledge and to construct what has been read interactively (Daneman & Hannon, 2001; Farr, Pritchard, & Smitten, 1990).

Some have raised concerns about the ability of multiple-choice test scores to reflect students’ reading skills mastery because the scores alone do not allow interpretation of the reading process (Daneman & Hannon, 2001; Farr et al., 1990; Logie 1995; Nevo, 1989; Phakiti, 2003; Tian, 2000). Test takers may not take the test in the way in which the test authors planned when designing it. For example, it has been argued that examinees do not need to read the texts upon which multiple-choice questions are based (Cohen, 1984; Keenan & Betjemann, 2006). Thus, a shift from focusing on test-takers’ scores to their reading processes and strategy use may provide insight into multiple-choice items that test reading comprehension.

Although different strategies with different testing formats are worth investigating, the present study focuses only on the strategies used to answer reading comprehension

multiple-choice questions because (a) there has been little, if any, research conducted on strategies regarding topic familiarity in the reading comprehension questions; and (b) multiple-choice is the most commonly used format that appears on high-stakes reading comprehension exams (Campbell, 1999; Phakiti, 2003).

Problem Statement

Current high-stakes reading comprehension tests are powerful assessments for determining and distinguishing test-takers' reading achievement. Concerning the use of multiple-choice questions to assess reading comprehension, researchers have argued that examinees' ability to select the correct answer from a list of distracters fails to measure examinees' actual reading ability, which threatens test validity. Researchers have maintained that examinees are more able to engage in a natural and interactive reading process if they are allowed to construct meaning by themselves (Campbell, 1999; Cohen & Upton, 2007).

Readers face additional challenges when faced with unfamiliar topics that restrict their topical knowledge activation. Therefore, it is necessary to investigate the multiple-choice test format by exploring the reading processes that test-takers employ to achieve answers based on different topics and by further focusing on the relationship between strategy uses and test scores. Without the evidence that multiple-choice questions adequately reflect examinees' reading ability, the validity of high-stakes multiple-choice reading comprehension tests is only theoretically assumed.

Purpose of the Study

The purpose of this study was to investigate the strategies that Mandarin-speaking students adopt when they read a text and then answer multiple-choice reading comprehension questions. The focus was on describing how students arrive at their answers when taking reading comprehension tests by asking them to verbalize their strategies.

Current reading research views reading as an interactive process in which readers build a personal understanding about the text, but unfamiliar topics restrict readers' activation of their background knowledge. The goal in this study was to explore students' reading strategies when dealing with familiar versus unfamiliar topics in the multiple-choice question test format. Also, strategy use was correlated with participants' scores on the reading test.

Significance of the Study

This study serves as a first formal attempt that I am aware of to investigate the strategies adopted by Mandarin-speaking students when they encounter familiar and unfamiliar topics in a multiple-choice format reading comprehension test. Although different task requirements and readers' language proficiency determine readers' strategy selection, there remains a critical lack of empirical studies that examine reading strategies in terms of topic familiarity, especially in the TOEFL iBT. Mandarin Chinese-speaking participants are an important population to study because they represent the largest group of students who took the TOEFL iBT test to gain admission to the University of Iowa. Additionally, Chinese-speaking population also represents the largest group of students in the U.S. who took the TOEFL iBT test. The findings have far-reaching implications for both language educators and test takers, and contribute to the field of reading strategies in general and test-taking strategies in particular.

Primary Research Questions

The research was designed to address the following questions:

1. For academic English as a second language multiple-choice reading comprehension tests, what comprehension strategies do Mandarin Chinese readers use when reading about familiar topics?
2. What comprehension strategies do native speaking Mandarin Chinese readers use when reading about unfamiliar topics?

3. How do the strategies compare when reading about familiar versus unfamiliar topics?
4. Do strategies differ between students with high and low scores on the reading test?

Definition of Terms

Comprehension strategies: The mental processes or behaviors that language learners consciously employ when accomplishing language tasks (Cohen & Upton, 2007; Erler & Finkbeiner, 2007).

Test validity: The degree to which correct inferences can be drawn based on results from an instrument; depends not only on the instrument itself but also on the evidence and theoretical rationales and characteristics of the group studied (Chapelle, Enright, & Jamieson, 2008).

TOEFL iBT: The latest version of the TOEFL, which intends to “measure the ability of non-native speakers of English to use and understand English as it is spoken, written and heard in academic settings” (Alderson, 2009, p. 621).

Prior knowledge: The “information, knowledge, emotion, experience and culture” that readers bring to the printed word (Brown, 2011, p. 299).

Multiple-choice questions: Test format in which test takers are asked to select one correct answer from multiple options.

CHAPTER 2

LITERATURE REVIEW

The purpose of this chapter is to provide a review of theories and research about the reading process, in both L1 and L2. This review provides relevant literature as well as the motivation for this study, which investigates different strategies used by Mandarin-speaking graduate students when they read familiar and unfamiliar topics on English reading comprehension tests with a multiple-choice question format.

This chapter is divided into four sections. The first section introduces reading theories, with a focus on the important role of readers. The second section pertains to the impact of readers' background knowledge, which is further divided into content schema and formal schema in reading comprehension. The third section discusses reading strategies, followed by strategy use according to readers' language proficiency and L1 pedagogical cultural backgrounds. The last section addresses the multiple-choice question format in L2 reading comprehension assessment, with a focus on test validity and multiple-choice test strategies.

The Nature of Reading

The view of reading comprehension as a hierarchical combination of several essential small components prevailed before 1985. It was believed that the successful mastery of subskills, such as readers' ability to recognize words and to understand the details of the text, would lead to success in both L1 and L2 reading comprehension. According to Johnston (1984), "the skills or component approach to reading comprehension is based upon the assumptions that comprehension can be analyzed into various discrete subprocesses, all of which are necessary for successful performance of mature reading" (p. 3). In this regard, mature reading comprehension was considered the interaction among subskills beginning with the knowledge of the alphabet, word recognition skills or phonological knowledge, the ability to group vocabulary into phrases

and sentences, and leading to an understanding of the meaning of the paragraph within the whole text (Blanton, 2004).

This subskills view of reading focused primarily on text decoding and echoed bottom-up reading approaches (Smith, 1986). Bottom-up reading researchers argued that readers comprehend by examining small units, then move to larger units, and a reader's major task is to decode the textual symbols that the writer intended to provide. Therefore, reading, in this case, was regarded as a one-way, single communication from the writer to the reader, and reading was seen merely as a receptive skill (Tian, 2000). Following this view, reading instruction at that time focused on teaching grammatical categories of individual words, sentence structure, and so on, because fluent decoding of these small and separate components would ultimately lead to successful comprehension (Orasanu & Penney, 1986).

However, the subskills' view of reading was challenged in the 1980s. Reading comprehension was newly defined as the "process [in which] readers construct a mental representation of the author's message, which includes both the information in the text and its interpretation by the reader" (Radojevic, 2006, p. 14). For example, Dechant (1991) argued that readers' background knowledge is critical in assisting readers to construct meaning from the text. Instead of viewing readers as passive decoders, researchers emphasized the role of readers as they actively engaged in the reading process by the knowledge they brought to the text. The next section focuses on the important role of the reader in reading comprehension.

The Important Role of the Reader

Although reading was once viewed simply as a receptive skill, contemporary reading theory views reading as a highly complicated and creative process in which readers use various resources to create meaning from the printed text (Bernhardt, 2005; Cooper, 1986; Grabe, 2009). According to Rumelhart (1985), reading, whether in L1 or L2, involves three key elements: the reader, the text, and the interaction between the

reader and the text. Reading comprehension depends on the information from the written text and the information that is retrieved from readers' background knowledge (Dechant, 1991). Therefore, reading comprehension is the process of "using the cues provided by the author and reader's prior knowledge to infer the author's intended meaning (Johnston, 1983, p. 9). In this view of newer reading models, readers actively construct meaning from the text by interpreting the information that the writer wants the reader to understand. Given that the "interaction between the reader and the text is the foundation of comprehension" (Cooper, 1986, p. 3), decoding skills were no longer the primary focus but were only one aspect of successful reading comprehension.

In recognizing the individual reader in the reading process, Bernhardt (1986) argued that the reader was one of the two primary variables in her multidimensional model, which contributed a valuable theoretical foundation to L2 reading comprehension (see Figure 1). The "reader-based" factors include intratextual perception (the reconciliation of each paragraph within a text), prior knowledge (the outside knowledge related to the text), and readers' metacognition (the extent to which readers are monitoring their reading). The "text-based" factors include word recognition (the semantic value of a word), phonemic/graphemic decoding (the match of pronunciation with the graphic part), and syntactic feature understanding (grammatical features). It can be inferred from the model that reading comprehension requires interaction between the linguistic elements in the text and the knowledge elements in the reader.

In this process, comprehension occurs when readers construct the text based on their knowledge of the reading task, their knowledge of the target language, and their knowledge about the world (Block, 1986, 1992; Erler & Finkbeiner, 2007; Grabe, 2009; Shapiro, 2004).

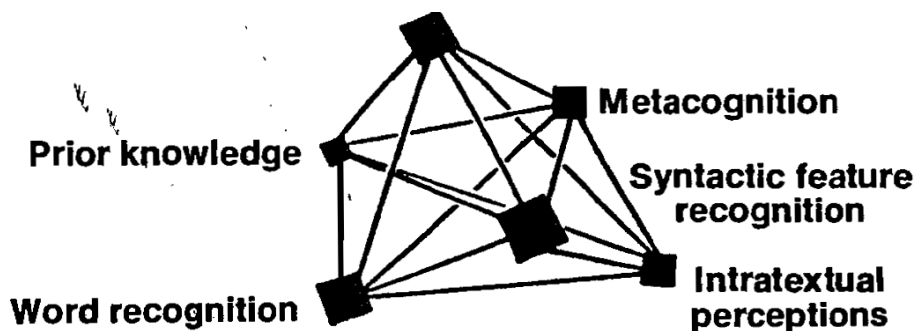


Figure 1. Bernhardt's (1986) constructivist model

Source: Bernhardt, E. B. (1986). Reading in the foreign language. In B. H. Wing (Ed.), *Listening, reading, and writing: Analysis and application* (pp. 93-115). Middlebury, VT: Northeast Conference on the Teaching of Foreign Languages.

Based on Bernhardt's and Rumelhart's theoretical models of reading, the communication between the reader and the text gained increasingly more interest in the reading field. As Kim (2010) argued, "a text by itself does not carry meaning, but rather guides readers in retrieving meaning based on their own prior knowledge" (p. 36). Thus, readers may differ in the meaning that each associates with a given word. With the widely accepted role of active readers, reading is no longer a series of skills that are sequential and hierarchical. However, it has been observed that readers are active learners who construct meaning by directing their own cognitive resources and prior knowledge to relate to the text (Garner, 1987; Logie, 1995). The interactive component of the reading process has been acknowledged by researchers in the L2 field (Bernhardt, 2005; Carrell, 1985; Grabe, 2009; Urquhart & Weir, 1998), and research is increasingly considering the variables of each individual reader, such as gender (Brantmeier, 2005; Oxford, 1993), language proficiency (Anderson, 1991; Huang, Chern & Lin, 2006; Phakiti, 2003), and sociocultural background (Singhal, 1998; Upton & Lee-Thompson, 2001).

Topic Familiarity in Reading Comprehension

With the notion that reading comprehension depends on information from the written text and information that is retrieved from readers' background knowledge,

readers' background knowledge in reading has received more research attention. As Brown (2001) noted, background knowledge is the "information, knowledge, emotion, experience and culture" that readers bring to the printed word (p. 299). A number of studies have provided empirical evidence to support the notion that topic familiarity can be an important factor affecting both L1 reading (Johnston, 1984; Recht & Leslie, 1988) and L2 reading (Brantmeier, 2005; Carrell, 1987; Chen & Graves, 1995; Droop & Verhoeven, 1998; Johnson, 1982; Krekeler, 2006; Malik, 1990). Moreover, several researchers have pointed out that this non-decoding variable may impact reading comprehension much more than readers' language proficiency (Hudson, 1988; Shapiro, 2004); as Ertmer and Finkbeiner (2007) argued, "the major difference between L1 and L2 reading is that L2 readers who are not familiar with content schema or do not process appropriate L2 sociocultural knowledge will have comprehension difficulties in that they cannot perceive the L2 text in a culturally authentic way" (p. 198). Thus, background knowledge facilitates reading in a more effective way and has a marked impact on reading comprehension.

As mentioned, the positive relationship between sufficient background knowledge and reading performance has been demonstrated in many L2 reading comprehension studies (Brantmeier, 2005; Hudson, 1988; Leiser, 2003). Background knowledge, or schema theory, is related to knowledge stored in readers' memories, and this theory explains how the new information is integrated with readers' previous knowledge, which facilitates reading comprehension (Alderson, 2000; Rumelhart, 1985). With the evidence that new information is learned and remembered the most when it is connected to related prior knowledge, activating the appropriate schema that fits the text is expected to be indispensable in the reading comprehension process. For example, Hudson (1988) pointed out that one of the L2 reading problems lies in the lack of activating the appropriate schema. With the wrong schema in mind, the reader will distort the text's meaning and find reading to be a difficult, even a laborious, task.

Schema theory is related to top-down reading processing, which allows readers to make predictions about the upcoming text and helps them to fill in the gaps when ideas are not explicitly stated in the text (Alderson, 2000; Hudson, 1988; Tsui, 2002). In this way, a schema assists readers to draw inferences and to guess meanings of those unclear or ambiguous messages from the text. Therefore, the appropriate schema compensates for readers' insufficient linguistic knowledge because those unknown messages can be inferred and decoded (Carrell & Wise, 1998; Johnston, 1984; Recht & Leslie, 1988). Thus, it may be possible that readers at a lower level of language proficiency could perform either as well as or better than readers at a higher level of language proficiency, given the correct background knowledge (Chen & Donin, 1997; Leiser, 2003).

To further explore the role of schema in ESL reading comprehension, Carrell (1987) distinguished two categories of schema. The first type of schema, content schema, is the reader's background knowledge associated with the text (e.g., information about African American culture). The second type of schema, formal schema, is the knowledge of discourse or rhetorical organization of the text (e.g., differences in the structures of stories and newspapers). Researchers have suggested that both schemata are necessary to fully understand a text or complete a testing task well (Singhal, 1998). The following literature review will focus on the influence of content schema and formal schema on L2 reading.

The Impact of Content Knowledge

Content knowledge is defined as the prior and relevant knowledge related to the content of a text. For example, a text may be based on the lives of frogs, and readers need to activate knowledge on the species of frogs. Content knowledge has attracted initial investigation from researchers in the realm of English as a first language (Johnston, 1984; Recht & Leslie, 1988). Motivated by those L1 studies, L2 researchers have also examined the role of content knowledge in reading comprehension. Johnson (1982) examined the impact that cultural knowledge possessed by 72 ESL students had on their

comprehension when reading passages about Halloween. The passages were divided into sections of familiar customs about the holiday and Halloween information that was assumed to be unfamiliar to the subjects. Using written recall and true-false questions, the author found that subjects recalled more information from familiar passages and distorted more information from unfamiliar ones. Therefore, the familiarity of cultural knowledge was effective in readers' comprehension.

Recht and Leslie (1988) conducted a study to examine whether prior knowledge or reading expertise had more influence on students' recall. Sixty-four junior high school students were selected and divided into a high L2 proficiency group that scored above 70% and a low L2 proficiency group that scored lower than 30% on the Scientific Research Associates (SRA) achievement comprehension test. Also, subjects completed 42 multiple-choice questions on their content knowledge test of baseball. For the study, subjects were required to read a passage about baseball games and provide verbal retellings, summarization, and non-verbal reenactment. Based on the results, the authors concluded that students with high background knowledge of baseball recalled more information, summarized more correct information from the text, and made more correct reenactments than did students with less knowledge. However, students with high proficiency and high content knowledge did not perform better than students with low proficiency and high knowledge. Therefore, Recht and Leslie (1988) concluded that "knowledge of a content domain is a powerful determinant of the amount and quality of information recalled, powerful enough for poor readers to compensate for their generally low reading ability" (p. 19).

In a similar study, Chen and Donin (1997) investigated whether content knowledge and language proficiency influenced students' reading comprehension. Thirty-six Chinese ESL learners were selected to read three passages about biology. Half of the participants were biology majors and the other half were engineering majors. Two of the three passages were in English and one was in Chinese. The subjects recalled the

information of the passage in their L1 for the Chinese text and recalled it in English for the two English texts. The results showed that participants with a higher level of background knowledge recalled more information on both the Chinese and the English texts, regardless of their language proficiency. There was no significant difference between the recall in L1 or L2. This result is consistent with the finding from Johnson's (1982) study, which suggested that sufficient content knowledge has a facilitative role in reading performance.

The concept of a background knowledge threshold, which suggested that readers must achieve a certain level of background knowledge for a schema to work, was demonstrated in the following studies. Recruiting 842 non-native English speakers, most of whom were about to matriculate as undergraduates at English-speaking universities, Clapham (1996) investigated the relationship between the language ability of Business and Social Sciences students taking reading comprehension questions on the IELTS test and students' ability to understand texts in and out of their disciplines. From these tests, students who scored below 60% were unable to understand texts even related to their own discipline, and students who scored 80% or above had little difficulty in comprehending the texts out of their field of study. Clapham concluded that background knowledge facilitated comprehension only when a minimum level had been reached. This finding fit the assumption of the two-threshold hypothesis, which argued that students with lower levels of proficiency could not take advantage of their background knowledge because they were too focused on vocabulary decoding, whereas high-level proficiency students could perform like native speakers who do not rely so heavily on their background knowledge. Thus, the scores of medium proficiency students were mostly affected by their background knowledge.

To further investigate the two-threshold effect, Krekeler (2006) recruited 400 international students from German universities, with the largest group of students from China. Participants were asked to complete two "Language for Specific Academic

Purposes” exams after reading two passages from a textbook for young speakers of German who were preparing for their school finals or for university study. The texts were subject specific, one related to business and the other to the measurement of velocity. From the results measured by a cloze test, the two-threshold hypothesis occurred only with the business topic. However, scores from the velocity passage suggested that students with medium levels of language proficiency benefited least from background knowledge, which did not substantiate the two-threshold effect. Krekeler explained that maybe there were too few very high and very low participants. In sum, these studies demonstrated the important influence of content knowledge on reading performance; however, some studies have brought counter opinions and argued about the extent to which schema can facilitate the reading process (Carrell, 1983; Hammadou, 1991; Hudson, 1988).

Some researchers have tried to explain why prior knowledge has not always been shown to occupy a facilitative role in reading performance in L2 reading. For example, if readers cannot correctly identify the meaning of the printed text due to their insufficient L2 proficiency, they will have difficulty in activating the correct schema that appropriately fits the text (Ridgway, 1997). Hudson (1988) suggested that the L2 reading problem lies in not being able to project “appropriate schemata” rather than in “failing to attempt to utilize schemata.” Shapiro (2004) echoed this statement by arguing that while correct prior knowledge assists learning, inaccurate prior knowledge can be more detrimental to learning than having no prior knowledge at all when reading. That is, students who had misconceptions about a piece of information performed less well than their counterparts who had no information. In addition, Carrell (1983) noted that ESL readers paid a good deal of attention to grammatical features, such as vocabulary and sentence structures, which left no room for schemata in their thinking processes. To summarize, previous studies have reached different conclusions about the influence of

background knowledge on reading comprehension, which provides motivation to investigate this effect further.

The Impact of Formal Knowledge

Formal knowledge in reading comprehension has been defined as “knowledge relative to the formal, rhetorical organizational structures of different types of texts” (Carrell, 1987, p. 461). Given that textual structure is produced in hierarchical linguistic devices, such as superordinate references like topic sentences, this type of knowledge provides readers with some expectations about text structures. Formal knowledge facilitates reading if readers are able to make accurate predictions about text types and genres (Lin, 2002; Radojevic, 2006; Resnick & Resnick, 1992; Singal, 1998). Kaplan (1966) observed that different languages have different writing patterns and discourse organization beyond the orthographical and grammatical differences. For example, Brown (2001) attributed to Kaplan (1966) the argument that “English discourse was schematically described as proceeding in a straight line, but Oriental written discourse in a spiraling line” (p. 337). English writing tends to be direct and straightforward, introducing the topic sentence at the beginning of the paragraph, whereas Chinese writing usually follows the traditional four sequences: introduction, elucidation of the theme, transition to another viewpoint, and summing up at the end. In this respect, Chinese writing is sometimes described as being “verbose, ornamental, and lacking in coherence” from a Western viewpoint (Singal, 1998). Therefore, Chinese readers may be confused and encounter comprehension problems if they are not familiar with the English language’s writing structure.

Studies have suggested that different types of text structure influence readers’ comprehension and recall. Stone (1985) examined whether language patterns in English, which differed from Spanish, would have a significant effect on ESL learners’ comprehension while reading English texts. Eighteen 5th-grade readers were assigned to an L1 Spanish-speaking group and the other 18 readers to an L2 English-speaking group.

Nine stories were developed with three different levels compared to participants' L1 language patterns: similar, moderately similar, and dissimilar. Students in each group read three stories, and comprehension measures included an oral reading, a retelling, and four comprehension questions for each story. Results showed that on the retelling measures, the lowest scores were found on stories that were the most dissimilar from the students' initial language and oral reading errors increased as language pattern similarity decreased. Although the comprehension scores did not vary significantly among formal familiarity, the results of retelling measures supported the assertion that texts that are not consistent with readers' expectations about language patterns can have disruptive effects on reading.

Such different patterns of organizational written structure were also found in a study by Eggington (1987). Eggington (1987) compared the two Korean rhetorical writing styles: the traditional rhetorical (non-linear) style, which is used more among local Korean scholars, and the U.S.-influenced (linear) rhetorical style, which is used more among U.S.-educated Korean scholars. The assumption is that U.S.-educated Korean scholars have been influenced in their L1 writing while studying in the U.S., and this disparity could have raised communication problems with local groups. By comparing Korean college students' memories of these two contrasting rhetorical forms, the author found a difference in recall. Korean college students studying in Korea recalled more information in the traditional non-linear framework text, whereas the newly returning Korean students from the U.S. remembered and displayed more information in a linear structure in their Korean writing, which was atypical to the traditional Korean rhetoric.

Because language structural dissimilarity may exist between any two languages, the differences in textual structure can lead to challenges for target language writers as well as readers. That is, beyond grammatical and lexical difficulties, it is very possible that target language learners will encounter additional problems when reading a text that is structurally unfamiliar to their first language's writing style. One way to manage this

problem caused by different formal knowledge has been suggested by Grabe and Stoller (2002), which is that instructors should understand students' L1 writing structure as much as possible and point out explicitly the structures of the texts to help students' reading processes. With this formal knowledge in hand, L2 readers will be able to comprehend L2 text more readily and in a culturally authentic way.

Strategies in Second Language Reading Comprehension

L2 researchers have identified reading strategies employed by readers with different language proficiency and in various reading contexts (Anderson, 1991; Bang & Zhao, 2007; Block, 1992; Brantmeier, 2005; Kim, 2010; Singhal, 2001). As Karbalaei (2010) described, reading strategies reveal the ways in which readers interact with the written text, such as how readers think and act when evaluating and planning their reading behavior, and how these behaviors contribute to text comprehension. L2 reading studies have moved from a focus on products (reading comprehension test scores) to a focus on reading processes and strategies. The term "strategy" has a variety of definitions. Rupp, Ferne, and Choi (2006) identified strategies as "conscious techniques and tactics deliberately employed by a reader for successful reading" (p. 447). Erler and Finkbeiner (2007) conceptualized strategies as "intentional actions chosen to facilitate reading at any level of processing" (p. 189). As these examples indicate, contemporary definitions of reading strategy are reasonably similar, focusing on activities performed by the reader to build meaning from the reading material.

Based on the importance of reading strategies, a plethora of studies have explored the strategies that L2 readers utilize to process a text. Pritchard (1990) conducted a study to investigate how readers' cultural knowledge affected their reading processes while they were reading one passage with which they were culturally familiar and another with which they were culturally unfamiliar. Thirty participants from the U.S. and from Palau, all 11th-grade students with high language proficiency, read a letter describing a funeral service in each of their cultures. Participants provided verbal reports about their strategy

uses while they read and retold the passage after reading. A total of 22 strategies were categorized under five groups: (a) developing awareness, (b) accepting ambiguity, (c) establishing intrasentential ties, (d) establishing intersentential ties, and (e) using background knowledge. Pritchard concluded that sufficient cultural knowledge significantly influenced the processing strategies that readers employed in strategies A and C for the culturally unfamiliar passage, and in D and E for the culturally familiar one. In addition, students recalled significantly more main units and reached higher comprehension levels from the culturally familiar text.

Another study examined the relationship between readers' gender, passage content, comprehension, and strategy use. Brantmeier (2000) recruited 78 native English readers of Spanish (29 male and 49 female) from an intermediate-level Spanish course to complete multiple-choice and recall comprehension tasks after reading two passages, one related to boxing and the other to housewives. From the results, males scored higher on the multiple-choice test and recall for the male-oriented passage and females scored higher on both tasks for the female-oriented passage. However, no significant difference between the global and local strategies that subjects used was found, which indicated that gender differences did not account for differences in strategy use. In addition, although the type and number of (global and local) strategy use remained almost the same for both passages, no positive correlation was found between subjects' global and local strategy use and their comprehension scores.

Although gender was not found to be a major factor, L1 influence was identified as a substantial factor for reading strategy use. Bang and Zhao (2007) conducted a research study about whether literacy skills transfer across languages, with particular attention to learners' word recognition and processing skills. That is, they examined how word recognition and processing skills of a learner's L1 may influence strategy selection in determining the meanings of L2 unknown words. By examining reading strategies of six advanced English learners from Korea and China, the authors confirmed that the L1

influenced students' L2 reading strategies. For example, because Korean is an alphabetic language and Chinese is an ideographic language, Korean participants generally used phonological processing strategies in reading, whereas Chinese students used visual-orthographic processing strategies. Without the evidence from the assumption that Korean ESL students would process English text more easily due to both Korean and English using an alphabetic writing system, the authors concluded that the comprehension level depended more on L2 language proficiency than on strategy transfer from L1. However, given that there were few participants in this study, and all of them were female doctoral students with very proficient reading ability, this study had a number of limitations.

More recently, Alsheikh (2011) conducted a case study to investigate the metacognitive reading strategies of three trilingual readers, with Hausa as the L1, French as the L2, and English as the L3. Participants were required to complete the Survey of Reading strategies (SORS), which included three broad categories of strategies: Global Reading strategies (GLOB), the generalized or global reading strategies; Problem Solving strategies (PROB), the localized or bottom-up decoding strategies; and Support Reading strategies (SUP), the use of support mechanisms such as dictionaries. From the findings, all three participants demonstrated a high awareness of strategy use, and the most frequently used strategies were PROB, followed by GLOB and SUP. In terms of different languages, multilingual readers employed more strategies in their L2 and L3 than in their L1. Namely, the number and types of strategies increased when readers encountered texts that proved to be more difficult.

In the studies described above, researchers investigated reading strategies through a variety of contexts and with diverse populations, who performed tasks with different text types, content, and difficulty. Even though there were many diverse aspects of each study, every study revealed important knowledge about L2 reading processes but with

contradictory results. A lack of strategy research has been reported when students read both familiar and unfamiliar texts.

Strategies with Different L2 Language Proficiency

Researchers have concluded that reading strategies are not used consistently across readers and there are clear distinctions between the strategies that proficient and less proficient readers use both in quantity and quality (Alderson, 2000; Almasi, 2003; Block, 1992; Hudson, 1988; Karbalaei, 2010; Kim, 2010; Koda, 2005; Saricoban, 2002; Singhal, 2001). As Hudson (1988) observed, “advanced level readers are in some way applying skills differently from the lower level readers, or rather, that the strategies which are chosen themselves change across levels of reading and language proficiency” (p. 198). In effect, successful readers are good strategy users; they are more aware of how they control their reading process and are able to verbalize their awareness. For example, good readers keep the meaning of the passage in mind while reading, use their background knowledge, and skip less important words (Block, 1992). Conversely, less successful readers are unaware of monitoring strategies while reading and overall apply strategies less effectively (Alderson, 2000). Because the differences in processing skills may result in a different level of comprehension, the following section will focus on studies of strategies with different language proficiency.

Anderson (1991) examined individual differences by comparing the reading strategies of 28 Spanish-speaking ESL students during academic reading and standardized test taking. Five broad categories of strategies were described: supervising strategies, support strategies, paraphrasing strategies, coherence-establishing strategies, and test-taking strategies. The results of Anderson’s qualitative and quantitative inquiries demonstrated that both proficient and non-proficient readers appeared to use the same kinds of strategies when answering questions. Thus, Anderson concluded that it is not enough to know what strategy to use, but that readers should be taught to “know how to use it successfully and orchestrate its use with other strategies” (p. 135) to become a

strategic reader. In addition, test items affect readers' responses and their interaction with the text.

Block (1992) conducted research on comprehension strategies of L1 and L2 learners of English. Participants were 25 college students, consisting of 16 proficient readers (eight L1 and eight L2) and nine non-proficient readers (three L1 and six L2). While reading the expository text, participants were asked to think aloud about two specific problems: a search for the referent and a vocabulary problem. Proficient readers identified the problems more frequently and applied background knowledge in assisting their understanding of the overall meaning of the passages. The less proficient readers focused more on local or word-level processing strategies. Block concluded that strategy use is a stable phenomenon that is not tied to a specific language, because "there is a regular process that operates similarly for native speakers of English and [proficient] second language readers" (p. 335). Thus, strategic knowledge seems to be more important than linguistic knowledge.

In a later study of strategy use in testing situations, Phakiti (2003) conducted research to report the relationship between test-takers' cognitive and metacognitive strategy usage and their test performance. The 384 participants from a Thai university were selected to take the gap-filling cloze and multiple-choice reading comprehension test, followed by a cognitive-metacognitive questionnaire about their thoughts while answering the questions. Among all of the participants, four highly successful and four less successful students were selected for retrospective interviews. Based on the results, there was a positive relationship between the test performance and the number of cognitive strategies used (Pearson's correlation $r = .4$) and metacognitive strategy use ($r = .5$), which indicated that readers with higher proficiency levels used more strategies. Moreover, metacognitive strategies were reported more by successful test-takers because they tended to be more aware of how and why they used specific strategies. Similar to

Anderson (1991), Phakiti also suggested that when strategies were employed was as important as which strategies were used.

Focusing on new forms of reading in the digital age, Huang et al. (2006) conducted a study to explore the online reading strategies of high- and low-proficiency EFL readers. The results of the TOEFL CBT test determined the proficiency level of the participants: 15 proficient students scored above 500, and 15 less proficient students scored less than 300. The strategies were categorized as follows: Global Strategy, Problem-Solving Strategy, Support Strategy, and Socio-Affective Strategy. The high-proficiency group used Support strategies (dictionary and translation) more than the low-proficiency group, whereas the low-proficiency group employed more Socio-affective strategies (listening to music). Huang et al. concluded that language proficiency resulted in different strategy use, with the high-proficiency group primarily using more Global or top-down strategies (keyword identification and prediction). The sequence of strategy use also varied according to language proficiency, as Global Strategies were used prior to Support Strategies for the high-proficiency group.

To conclude, some researchers have argued that there is a relationship between the types of reading strategies that readers use and their proficiency level (Block, 1992; Huang et al., 2006). That is, proficient readers tend to use more top-down than bottom-up reading strategies. Top-down strategies involved identifying main ideas in the text, predicting content by using their background knowledge, rereading and monitoring their reading processes, and making predictions (Bang & Zhao, 2007). However, Brantmeier (2000) showed that the use of global strategies led to both successful and unsuccessful reading comprehension. Furthermore, some studies did not connect the type of strategy used to successful comprehension. For this reason, studies about L2 processing strategies are needed that investigate the correlation between type of strategy and performance.

Strategies with Readers' L1 Pedagogical Cultural Background

In addition to their language proficiency, readers' L1 cultural literacy backgrounds may affect their L2 reading processes and strategy preferences. That is to say, language learners from different pedagogical cultures are likely to employ dissimilar reading strategies when comprehending the same text due to their varied educational backgrounds and experiences in literacy learning (Abbott, 2006; Block, 1986; Parry, 1996). Erler and Finkbeiner (2007) supported this observation by stating that "L2 reading cannot be separated from the social, cultural, institutional, and personal practices of L2 readers" (p. 198). With a possible contribution of readers' L1 literacy experiences, studies have focused on how L2 reading strategies interact with L1 cultural pedagogical backgrounds.

Research has demonstrated that different reading processes are employed by readers from different pedagogical cultural backgrounds. Parry (1993) conducted research about Chinese and Nigerian readers' strategies in reading English texts. The results showed that Chinese students focused on the details of language and preferred bottom-up reading methods, whereas Nigerian students were more interested in comprehending English in broad concepts and used more top-down reading methods. Parry concluded that different cultural groups used different reading strategies, which were related to readers' language and literacy backgrounds.

Abbott (2006) collected verbal reports from 15 ESL participants after they finished an English reading assessment. Seven Arabic and eight Mandarin-speaking intermediate language-ability students participated in the study. The differential item functioning method (DIF) was used to investigate whether the participants from different populations performed the same in their reading strategies selection. The results showed that Mandarin speakers preferred bottom-up strategies, such as focusing on lexical items and matching key vocabulary in the text. In contrast, Arabic participants used more top-down strategies, for example, skimming for gist and drawing inferences based on the information of the text. Abbott concluded that learners from different pedagogical

cultural and geographical backgrounds favored distinct reading strategies when approaching the same reading material.

Kohn (1992) hypothesized that Chinese-speaking learners preferred bottom-up strategies because K-12 Chinese teachers encourage students to (a) read slowly and take care that they know each word as they go; (b) vocalize or voice the reading, either loudly or silently; (c) reread difficult sentences until they are understood; (d) look up definitions of all unknown words in a dictionary; and (e) analyze complex structures carefully (p. 121). The way they were taught to approach reading may explain Chinese EFL students' tendency to use bottom-up reading strategies. As a result of this teaching approach, Chinese EFL learners may attend highly to details and encounter problems with integrating the text as a whole compared to learners who are more familiar with using global strategies.

Some pedagogical cultural and educational factors have been shown to influence strategy preferences, which supports the contention that the approach and comprehension of L2 reading depends on readers' L1 sociocultural and educational backgrounds (Upton & Lee-Thompson, 2001). For this reason, readers' pedagogical cultural background should be one of the variables in strategy investigation.

Second Language Reading Comprehension Assessment

To determine whether and to what extent readers understand what they read, teachers often rely on students' results on reading comprehension assessments. Before the mid-19th century, L1 or L2 reading comprehension was measured by oral reading, which was later viewed as a poor assessment instrument due to its lack of connection to reading comprehension (Moore, 1983). Next, readers were assessed using recall tasks, which asked test-takers to write down or speak everything they remembered from the text. However, this technique seemed to measure a students' memory as well as their reading proficiency. Later, other types of assessments emerged that required students to read short passages and answer questions based on these passages (Tian, 2000), such as cloze tests,

multiple-choice tests, true-false tests, short-essay questions, and so on. The following sections introduce the validity of reading comprehension assessment, focusing on multiple-choice formats and related strategy use of this format in reading comprehension tests.

Validity and Reading Comprehension Tests

Reading comprehension assessments have evolved with the theories of reading comprehension, from a single decoding process to an interactive process in which readers actively construct meaning from printed texts. Before the 1980s, traditional reading comprehension tests, in general, viewed the reader as simply a decoder; therefore, reading comprehension tests at that time focused primarily on measuring isolated subskills, such as knowledge of word recognition, punctuation, or vocabulary spelling (Tian, 2000). In addition, when compared to the texts that students were reading in the classroom, passages in traditional reading comprehension tests tended to contain more information and less structure because they were designed only for testing purposes (Campbell, 1999).

Given that a number of researchers have argued that reading tests as a mastery of isolated subskills have not kept pace with current reading theories (Abbott, 2006; Alderson, 2000; Cohen & Upton, 2007; Logie, 1995; Valencia & Pearson, 1987), there has also been a change in reading comprehension assessment. As a result, reading comprehension tests have evolved from evaluating students' reading subskills to the realm of considering students' performance as dependent on "characteristics of the text, the nature of the task, and the context as well as the person's reading abilities and prior knowledge" (Johnston, 1984, p. 21). This shift assumes that skilled readers are engaging in reading processes while they actively and strategically examine their comprehension and make adjustments when necessary.

With the calls for improvement in assessing reading skills, the assessment of reading comprehension has become more compatible with what society might expect

students to learn in the real world. The top priority in selecting and using any assessment should be whether it helps students' reading skills and the instruction of reading in language classrooms (Campbell, 1999). In the past, the product of a reading assessment (i.e., usually a numerical score) had more impact than the process, as test-takers' abilities were determined by a single reading comprehension score. However, scores cannot reflect the processes of how the students construct meaning during test taking. As Logie (1995) argued, the products are far more meaningful if they are accompanied by information about readers' in-process thinking and reading strategies, which elucidate students' reasons for the responses they make. In this regard, rather than just focusing on the score examinees receive, there has been a growing trend to pay more attention to the reading processes and strategies that readers employ.

The modifications of reading assessments show a major concern for test validity. According to Bachman and Palmer (1996), validity is related to the "meaningfulness and appropriateness of the interpretations that people made on the basis of test scores" (p. 21). Test validity should be judged by whether the test score truly reflects examinees' reading ability and to what extent the processes elicited during test-taking situations are comparable to the processes in normal reading situations (Bachman & Palmer, 1996; Campbell, 1999; Chappelle et al., 2008; Cohen, 2011). For this reason, validity is regarded as the most important consideration of test evaluation and application because a fundamental component of a good test is that it must adequately sample the domain to be measured (Chappell et al., 2008; Daneman & Hannon, 2001; Logie, 1995; Messick, 1994).

However, many testing authorities have questioned the validity of reading comprehension assessments, suggesting that the scores fail to describe examinees' actual reading ability. For example, Nevo (1989) argued that there was a discrepancy between test planners' assumptions about what they intended to test and the actual processes that respondents engaged in during test taking. Namely, examinees may not read the test

instructions or answer the questions in a way that test examiners planned when they were designing the tests. Under such circumstances, many tests have been criticized for allowing correct answers to be reached without examinees actually understanding the text or using any judgment activity in selecting the correct responses (Cohen, 1998, 2011). If reading skills that are intended to be assessed in a reading comprehension test do not adequately represent the construct of reading comprehension demonstrated by test takers' behaviors, it is difficult to draw firm conclusions about an examinee's ability using a single test score.

Multiple-choice Format in Reading

Comprehension Assessment

Presently, different test formats are used to assess reading comprehension, for example, multiple-choice, open-ended, cloze, true/false, recall summary, gap-filling, and even the use of portfolios (Campbell, 1999; Tian, 2000). Among these formats, multiple-choice questions, which provide examinees with relatively short passages followed by several questions about that passage and require test-takers to select a correct answer among other distractors, seem to be the most commonly used yet often the most controversial. As Johnston (1984) observed, multiple-choice items are “probably the most researched, most maligned, most difficult to construct, most abused, yet most functional of all items” (p. 59). The proliferation of multiple-choice as the primary format in large-scale exams can be attributed to several reasons. The format is valued for its practicality and scoring efficiency, with a quantifiable number that allows easy comparison across test-takers, especially in high-stakes exams. In addition, multiple-choice responses are machine-scoreable and objective, which “minimize the disadvantages inherent in assessment procedures that required subjective rating” (Campbell, 1999, p. 35).

However, researchers have identified problems with these conveniences. A major problem is that multiple-choice questions may engage test-takers in a guessing game,

because the correct answers can be reached without reading the passage, let alone comprehending its meaning (Cohen, 1984; Logie, 1995; Ozuru et al., 2007; Tian, 2000). Proficient readers may select wrong answers for the right reasons whereas poor readers may select right answers for the wrong reasons; as a result, the multiple-choice questions may reveal little about test-takers' comprehension process, and its diagnostic value may be questionable.

With the problem mentioned above in mind, studies have been conducted to investigate the validity of multiple-choice formats by comparing multiple-choice and constructed-response questions, such as free-response and recall (Bridgeman & Rock, 1993; Campbell, 1999; Rodriguez, 2003; Van den Bergh, 1990). Given that different comprehension assessment tasks may not be testing the same comprehension ability (Andreassen & Braten, 2010; Brantmeier, 2000), studies have been focused on the various formats used in reading comprehension tests and whether they elicit the same test-taking processes. Some researchers have argued that little or no difference exists between multiple-choice and constructed-response formats. For instance, Bridgeman and Rock (1993) determined that multiple-choice and open-ended versions of the Graduate Record Examination (GRE) test measured basically the same construct by running a factor analysis of test scores. This same result was found in Van den Bergh's (1990) study, which demonstrated that "students seem to construct their answers to multiple-choice items to the same degree as when they answer open-ended reading comprehension items" (p. 10). Ozuru et al. (2007) further argued that the comparability between multiple-choice questions and open-ended items depended on the quality of the distractor items. These studies supported the comparability of multiple-choice questions and free constructed-response questions.

However, different formats yield different degrees of complexity for test takers, and students' ability to demonstrate their comprehension is affected by the tasks on which they are tested. For example, Wolf (1993) found significant differences between

multiple-choice and constructed-response formats and suggested the need for further studies to directly compare subjects' responses to examine format difficulty. In a similar study, low-proficiency students performed better on multiple-choice items than on open-ended and cloze items (Shohamy, 1984). Some researchers cautioned against that higher order thinking and in-depth reading skills, which should be the goals of reading instruction, are not assessed in some multiple-choice questions (Cohen, 1984; Resnick & Resnick, 1992). Thus, many have advocated the use of various testing formats to assess reading comprehension to obtain a more thorough evaluation (Andreassen & Braten, 2010; Campbell, 1999; Powell, 1988; Wolf, 1993; Tian, 2000).

Multiple-choice Format and Topic Familiarity

Some studies have suggested that multiple-choice tests prevent students from applying their background knowledge to assist reading. The controversy surrounding the multiple-choice format stems from the restrictions imposed on readers that a single correct answer must be determined among a number of incorrect options. This limitation concerns many reading experts because reading is currently regarded as a constructive process, which is dependent on the interaction between the reader and the text. Therefore, expecting all test-takers to reach the same interpretation about a text seems to contradict the current view of reading (Farr et al., 1990; Radojevic, 2006; Tian, 2000).

According to Farr et al. (1990), there is an ongoing issue about the readers' "ownership" of a text, and it seems that multiple-choice tests "reinforce the notion that there is inherent in text a correct meaning that readers must determine" (p. 209). Valencia and Pearson (1987) also argued that the diversity between test-takers' prior knowledge and experiences "invite many possible inferences to fit a test or question" (p. 731). As readers constantly build a personal understanding of the meaning of a text, the multiple-choice question format does not account for readers' background knowledge or allow readers to interpret text creatively (Alderson, 2000; Campbell, 1999; Farr & Carey, 1986; Farr et al., 1990; Tian, 2000). Therefore, to perform successfully on a

multiple-choice test, examinees need to suppress the use of their background knowledge and try to select a predetermined correct answer, even though other choices may be acceptable from different perspectives.

As studies have shown that topic familiarity facilitates reading comprehension, researchers have been curious about the extent to which readers activate their background knowledge when answering multiple-choice assessments. To investigate the influence of background knowledge on multiple-choice reading tests, Hudson (1988) conducted a study to examine the role played by background knowledge in L2 reading assessment by ESL students. A total of 93 ESL students studying in the U.S. were divided into two groups, with one group being taught relevant knowledge before the reading assessment and the other group receiving no instruction. After reading, students were required to take a multiple-choice test with 10 questions. The results showed that students with low or intermediate language proficiency scored higher after they received prior knowledge instruction. In contrast, the effect was not significant among proficient students. Hudson (1988) concluded that differences existed in the abilities to activate schemata from printed words between readers with different levels of language proficiency. Thus, activating correct background knowledge seems helpful in improving reading comprehension performance with multiple-choice format questions.

Peretz and Shoham (1990) conducted a study to determine whether ESL students' reading comprehension performance was modified if the reading topic was related to the field of study with which they were familiar. One hundred and seventy-seven Israeli students were selected from humanities and sciences fields; students were asked to read two articles, one related to the humanities and the other to the sciences. Both articles were controlled for difficulty and structure. Based on the results of 14 multiple-choice questions, there was a non-significant relationship between topic familiarity and students' performance because the science students scored higher on topics with which they were unfamiliar. Peretz and Shoham provided the explanation that science students might have

higher language proficiency that influenced the scores more than their domain-specific knowledge.

Ozuru et al. (2007) examined the effects of passage availability (the influence of background knowledge) and different reading comprehension test formats (multiple-choice and open-ended questions). Forty-one college students read a text related to their background knowledge and answered 12 open-ended questions and 13 multiple-choice questions. Twenty participants were allowed to look at the passage while answering the questions while the others were not. The results supported the hypothesis that “making the text available while answering comprehension questions tends to reduce the impact of prior domain knowledge” (p. 421). That is, if the text is available, readers are inclined to use textual information to find clues in comprehension question answering instead of relying on their background knowledge. Ozuru et al. (2007) also concluded that the performance on multiple-choice questions was highly correlated with performance on open-ended questions when the text was not available, but not in the with-text situation.

Multiple-choice Items and Strategy Use

Under the assumption that different contexts determine readers’ strategy selection (Singhal, 2001; Phakiti, 2003), strategies should be considered in relation to reading purposes because they are applied distinctively in each specific situation. Accordingly, studies have compared the strategies used in normal reading conditions versus those in high-stakes testing situations and concluded that some strategies are specific to test-taking situations (Alderson, 1991; Daneman & Hannon, 2001; Farr et al., 1990; Phakiti, 2003; Shohamy, 1984). According to Cohen and Upton (2007), test-taking strategies are defined as “test-taking processes which the respondents have selected and which they are conscious of, at least to some degree” (p. 211). Furthermore, Rupp et al. (2006) argued that different testing formats resulted in different strategy use. Given that

the multiple-choice format is the major focus of this study, the following review includes only studies related to multiple-choice reading tests.

One of the major concerns regarding multiple-choice questions is the use of test-wiseness strategies (Dolly & Williams, 1986; Hill & Larsen, 2000; Wolf, 1993). Test-wiseness has been defined as the “ability to use test-taking strategies to select the correct response in multiple-choice tests, without necessarily knowing the content or using the skill that is being tested” (Allan, 1992, p. 101). As mentioned, researchers have expressed concern that examinees’ ability to select the correct answer from a list of distracters has little connection to their reading comprehension. Researchers are not only concerned about the inadequacy of truly reflecting test-takers’ reading abilities, but they also fear that the common use of multiple-choice tests has led to the development of reading strategies that are detrimental to actual reading experiences. For example, multiple-choice questions engage test takers in a guessing game (Farr et al., 1990). In addition, researchers have pointed out that many multiple-choice questions can be answered without reading the passage related to the questions (Cohen, 1984; Tian, 2000).

Cohen (1984) reported the results of several studies on strategy use when EFL students took multiple-choice question tests. From the students’ self-reports, several strategies were identified: searching for corresponding questions, matching the passage and item stems and alternatives, locating words that were repeated in the sentence or key word association, among others. Concerned about the ability of test scores to represent examinees’ reading performance, Cohen suggested that researchers should make an effort to get a closer fit between “how test constructors intended for their tests to be taken and how respondents actually take them” (p. 79). Such a goal can be reached by modifications in the test format and training students in test-taking situations by exploring their question-answering rationale.

Farr et al. (1990) studied 26 Midwestern college seniors, asking them to describe their thoughts while taking a multiple-choice question test. This study was intended to

investigate what strategies examinees used in a typical test-taking situation. The results were divided into three groups: overall strategies (how much the passage was read before examinees answered the questions), reading strategies (how the examinees read the passage), and test-taking strategies (how examinees moved back and forth between the passage and the questions). The findings suggested that students viewed questions and passages as one interrelated task rather than only concentrating on the questions. Therefore, contrary to Cohen's (1984) view, Farr et al. (1990) concluded that multiple-choice tests provided a "reasonable estimate of readers' overall reading ability" (p. 224) and also supported the construct validity of multiple-choice reading comprehension questions. However, the study was limited by the choice of college seniors as subjects, because mature readers tend to use different strategies than younger and less proficient learners.

Similar to the argument of Farr et al. (1990), Daneman and Hannon (2001) conducted a study using the working memory theory to investigate the construct validity of multiple-choice reading comprehension tests, such as the Scholastic Assessment Test (SAT). The study was conducted with 48 college students in Canada who were native speakers of English. From the results, test takers received the poorest level of performance (35%) if they did not read the passage at all; they achieved an intermediate level (63%) if they used the questions to search for the answers without global reading of the passage; they achieved the highest level of performance (71%) if they read the entire passage in advance and reread portions of the passage to answer the questions. Thus, Daneman and Hannon (2001) suggested that test takers read the passage as much as possible to enhance global comprehension in order to achieve higher scores. Moreover, Daneman and Hannon argued that the SAT test maintains test validity because examinees still need to demonstrate complex verbal skills if they do not read the passage.

Cohen and Upton (2007) examined test-taking strategies on the *LanguEdge Courseware* (2002) material, which was developed to familiarize future respondents with

the TOEFL iBT exam. The focus of this study was on the strategies used when test-takers responded to single-selection multiple-choice items (Basic Comprehension tasks, Inferencing tasks) and the new multiple-selection multiple-choice items (Reading to Learn tasks). Thirty-two high-intermediate to advanced participants whose native languages were Chinese, Japanese, Korean, and others were selected for this study. The time was limited to 25 minutes, as in a real test situation. Due to time and cost constraints, only 13 responses to the test were transcribed, and one task from each of the three broad categories was analyzed.

The most frequently used strategies in the three broad categories were as follows. For the Basic Comprehension-vocabulary item type, strategies such as jumping immediately to the word before looking at the options or reading a portion of the passage carefully were most frequently used. The Basic Inference item, which was regarded as the most difficult type, was intended to “measure examinees’ ability to comprehend an argument or an idea that is strongly implied but not explicitly stated in the text” (Cohen & Upton, 2007, p. 228). Strategies such as returning to the passage to search for clues or discarding and selecting answers based on the paragraph/overall passage’s meaning were commonly used. The Reading to Learn-prose summary was designed to “measure examinees’ ability to understand the major ideas and relative importance of information in a text” (p. 230). Participants used the following strategies: reading the options first before going back to the passage, rereading the question, paraphrasing the question, and rereading the portion of the passage again carefully. The authors demonstrated an array of strategies that were tailored to specific types of multiple-choice questions. The authors concluded that the new TOEFL reading test evaluates test-takers’ ability to use a fairly consistent combination of basic academic reading and test-taking skills to accomplish a variety of academic-like reading tasks.

Conclusion

The literature reviewed in this chapter provides the framework and motivation for this study. The nature of reading comprehension, readers' content and formal familiarity with a text, reading strategies employed by readers with different language proficiency and cultural backgrounds, and multiple-choice format reading comprehension assessments are all important aspects that helped to frame the research questions in this study.

The review brings to light some gaps associated with these studies. First, given the notion that what readers know affects what they understand, researchers have been interested in investigating the role of topical knowledge in reading comprehension. However, very few existing L2 studies have developed a thorough topical knowledge assessment to measure participants' topical knowledge. Instead, participants' topical knowledge is only assumed by the researcher according to their school major or cultural background.

Second, test-taking strategy studies have been increasingly important in helping to construct test validity by providing descriptions of how examinees reach their test responses. Accordingly, a multiple-choice format reading test cannot be claimed as valid unless it presents evidence that it adequately reflects examinees' reading ability. With many studies focusing on reading strategies, there seems to be a lack of studies focusing on test-taking strategies. Thus, a closer investigation of test-taking strategy use and reading assessment language proficiency is necessary.

Moreover, as Singhal (2001) and Phakiti (2003) argued, different task requirements and reading texts determine readers' strategy selection. Researchers have investigated reading strategies through a variety of contexts and with diverse populations who performed tasks related to different text types, content, and levels of difficulty. Unfortunately, based on the literature review, very few, if any, studies in the reading literature have examined the strategies that readers adopt when they are faced with

familiar and unfamiliar texts, especially in high-stakes multiple-choice exams. This gap suggests a need for research on strategies, given this specific condition. This study, which examines the strategy employment of students when confronted with familiar versus unfamiliar topics in a multiple-choice format reading test, may bridge the gaps and provide insights into reading comprehension test taking.

CHAPTER 3

METHODOLOGY

The main purpose of this research study was to compare the strategy use of Mandarin Chinese-speaking graduate students when confronted with familiar versus unfamiliar topics in a multiple-choice format reading test based on TOEFL iBT reading comprehension practice questions. The focus was on describing students' strategy use when they were taking a reading comprehension test by asking students to verbalize their thoughts. These strategy use reports were further examined with respect to participants' comprehension test scores. Both qualitative and quantitative analyses were used in this study.

Research Design

The mixed methods design has been defined as the “combination of qualitative and quantitative approaches in the methodology of a study” (Tashakkori & Teddlie, 1998, p. ix). Because qualitative research and quantitative research design provide different perspectives and each has its limitations, the limitations can be compensated by the strengths of the other method when the two methods are integrated. For example, participants' direct voices cannot be considered in quantitative research. In contrast, qualitative research often is regarded as deficient because of the limited number of participants studied. For this reason, mixed methods approaches have gained more popularity in research designs.

Mixed methods research design provides more comprehensive evidence than either qualitative or quantitative design alone. Using both types of methods broadens the research analysis and therefore strengthens the results. The results can lead to more powerful generalization without being restricted to the types of data collection and analysis typically associated with qualitative or quantitative research (Creswell & Plano Clark, 2011).

In order to answer the research questions that guided this study, I chose the Exploratory Design (Creswell & Plano Clark, 2011) and modified it to fit the purpose of the current study. This design is defined by exploring a research context with qualitative data and then measuring it with quantitative analysis. Thus, the quantitative phase is dependent on, and used to validate, the results of the previous qualitative phase. In the final discussion, specific qualitative findings were noted and interpreted with the quantitative results.

This present study started with a quantitative data collection and analysis, which were used to select participants who appropriately fit the condition of this study. First, I approached all 30 individuals who responded to the study invitation. The responding individuals were screened based on their scores from their topical knowledge vocabulary assessment and two reading comprehension tests. After the scores were calculated, ten individuals were removed because they did not fit the conditions of this study. In phase two, think-aloud protocols and one-on-one interview responses of the 20 remaining participants were collected and analyzed. In the third phase, the qualitative information obtained from phase two was submitted for statistical analysis. Finally, the qualitative and quantitative results were combined and interpreted together, with a closer examination of four selected participants as case studies. These four case studies were chosen due to their extremely high or low reading comprehension scores.

Since this study was designed and conducted with the qualitative phase being prioritized, this investigation took the form of “a marginally mixed” study (Teddlie & Tashakkori, 2003). This form of mixed methods study combines both qualitative and quantitative research, but emphasizes one research method more than the other. In this study, quantitative methods were used to elaborate on the qualitative data results obtained in the second phase, which were the focus of the study. According to Greene et al.(1989), the marginally mixed method helps one method to clarify the results of the other method.

This present study combined qualitative and quantitative analyses and interpreted the results based on the integration of data analysis.

The following sections describe the research questions, the participants who were recruited, the material and instruments that were used to elicit data on responses to familiar and unfamiliar topics in multiple-choice questions, the data collection procedures, and the data analyses used in this study.

Research Questions

The research was designed to address the following questions:

1. For academic English as a second language multiple-choice reading comprehension tests, what comprehension strategies do Mandarin Chinese readers use when reading about familiar topics?
2. What comprehension strategies do native speaking Mandarin Chinese readers use when reading about unfamiliar topics?
3. How do the strategies compare when reading about familiar versus unfamiliar topics?
4. Do strategies differ between students with high and low scores on the reading test?

Participants

Graduate students from Taiwan and China studying at the University of Iowa were invited to participate in this study. A total of 30 graduate students responded to the invitation (see Appendix A) and completed their participation. However, ten participants were removed from the data analysis because they did not fit the conditions of the study. Specifically, the data for four students were eliminated because they scored similarly on all of the content areas, and the data for six students were discarded because they scored neither high nor low on their reading comprehension tests. The total number of remaining participants was 20, and their characteristics are shown in Table 1.

Table 1. Participant Information

Participant ID	Gender	Country	Degree	Major	Semester enrolled
3	Female	Taiwan	Master's	Economics	4
4	Female	Taiwan	Doctorate	Foreign Language Acquisition Research and Education	6
5	Female	Taiwan	Doctorate	Accounting	1
7	Male	China	Doctorate	Electrical Engineering	3
8	Female	Taiwan	Master's	Foreign Language & ESL	3
9	Male	China	Doctorate	Computer Science	4
10	Male	Taiwan	Doctorate	Electrical Engineering	2
12	Female	China	Master's	Chinese Department	3
13	Female	Taiwan	Doctorate	Economics	2
14	Female	China	Master's	Business Administration	3
16	Female	China	Master's	Law	2
17	Female	Taiwan	Doctorate	Foreign Language & ESL	2
18	Male	China	Doctorate	Electrical Engineering	1
21	Male	Taiwan	Master's	Law	3
23	Female	China	Master's	Law	1
25	Female	Taiwan	Doctorate	Mechanical Engineering	3
26	Male	Taiwan	Master's	MBA	4
27	Female	China	Master's	Chinese Department	2
28	Female	China	Doctorate	Foreign Language & ESL	4
29	Female	China	Doctorate	Foreign Language Acquisition Research and Education	4

The participants, 12 females and eight males from both China and Taiwan, were from four academic disciplines: business, law, language teaching, and engineering. In terms of their degrees, nine out of 20 were master's students, and 11 were pursuing their doctorate degrees at the time of the study. Also, participants' semester enrolled in the University of Iowa is provided. In terms of their experience with the TOEFL iBT test, most of the doctoral students indicated not having such experience because the iBT test was introduced after they had applied for admission to universities in the U.S.

Materials

Materials used for this study were selected from the book, *Comprehensive TOEFL iBT Reading 2007-2009*, published by Harvard Press (2006) and the book, *TOEFL iBT*

Reading 120, published by Jinni Publishing Corporation in Taiwan (permission letters are included in Appendices B and C). These two books contain preparation tests for TOEFL iBT test-takers, and the items resemble the actual TOEFL iBT reading section. The actual TOEFL iBT reading section has from three to five texts with different topics, and each text is followed by 12 or 13 multiple-choice response items; test takers have from 60 to 100 minutes (20 minutes for each text) to answer questions related to the texts. The TOEFL reading test is designed to simulate the types of academic reading tasks that students are expected to engage in at the university level.

According to the Educational Testing Service (ETS, 2003), the reading section of the new version of the TOEFL iBT measures three broad categories of reading skills: basic comprehension, inferencing, and reading to learn. Because this study uses a simulation of the TOEFL tests, it measures participants' skills along similar lines. However, this study focuses only on the basic comprehension and inferencing questions because the reading to learn questions are different from the standard one-answer multiple-choice format. For the reading to learn questions, test-takers are required to select three answers out of six options provided. The two academic categories of reading skills focused on in this study are illustrated from the TOEFL iBT practice book in Table 2.

Each text in the practice tests has approximately 600-700 words, which is similar to the authentic TOEFL iBT reading test. There are normally 12 to 14 multiple-choice questions following each text, asking test-takers to demonstrate their understanding of the text. Given that this study will focus only on the basic comprehension and inferencing questions, the number of items will be limited to 10 each, with the elimination of the reading to learn questions. The reading tasks are included in Appendices D and E.

Table 2. TOEFL iBT Questions and Examples

Reading Tasks	Definition	Sample Item from Business Area
Basic Comprehension	Focus on the “ability to understand important information in a text based on the lexical, syntactic, and semantic content of the text” (ETS, 2003, p. 4)	Q 1: The word “rationing” in the passage is closest in meaning to? Q 2: Based on the information in paragraph 1, which of the following is true of the demand curve?
Inferencing	Focus on “sentence-level information” and “abilities related to connecting information and recognizing the organization and purpose of the text” (ETS, 2003, p. 25).	Q 1: What can be inferred from paragraph 2? Q 2: Look at the four squares that indicate where the following sentence could be added to the passage.

Data Collection Instruments

Three data collection instruments were developed and used for analysis in this study: a topical knowledge vocabulary assessment, a think-aloud protocol, and an interview. Each is introduced in detail in the following sections.

Topical Knowledge Vocabulary Assessment

In research on topic familiarity, a number of measures and operations have been used to assess readers’ topical knowledge, for example, multiple-choice questions (Carrell & Wise, 1998; Johnston, 1984), vocabulary assessments (Huang, 2010; Valencia, Stallman, Commeyras, Pearson, & Hartman, 1991), self-reports (Lin, 2002), and interviews (Valencia et al., 1991). In operationalizing topic knowledge, I drew heavily on guidance from Huang (2010) in developing a process for assessing participants’ topical knowledge in this study. Five phases were used in developing the assessment.

Phase 1: Assessment Identification

After considering all of the possible measurement approaches described above, I chose both vocabulary assessment and interview to assess test-takers’ topical knowledge.

Vocabulary assessment in this study required participants to write definitions of a word from the content of a specific topic. Vocabulary knowledge has been regarded as an appropriate evaluation for readers' knowledge of a given topic (Huang, 2010; Johnson & Pearson, 1984; Valencia et al., 1991). Word selection was based on the key concepts related to each topic chosen by content advisors. It was assumed that all presented vocabulary could be answered without reference to the text by someone who knew the topic.

Because readers' topical knowledge could not be measured by a single assessment, the data needed to be triangulated. Thus, interviews were conducted with each participant. As Kvale (1996) argued, an interview is one of the most powerful ways to understand the world from the subjects' points of view, to reveal the meaning of people's experiences, and to uncover their lived world. In this study, in addition to the qualitative interviews on reading patterns, self-reports of topical knowledge were solicited. Participants were asked to report their familiarity with the specific topic on a scale of 1-10, with 1 being *most unfamiliar* and 10 being *most familiar*, to confirm the results of the topical knowledge assessment.

Phase 2: Content Area Search

After deciding to use a vocabulary knowledge test and an interview response, I selected the specific reading materials from the TOEFL iBT reading practice books published by Harvard Press (2007) and the book *TOEFL iBT Reading 120* published by Jinni Publishing Corporation in Taiwan. I was interested in assembling a broad range of topics varying widely with respect to topical knowledge. From a range of differing texts and topics, four articles discussing separate ideas were selected. The four texts included the topics of law (Civil Law), business (Law of Demand), language teaching (Reflection in Teaching), and engineering (Electricity from Wind). Most important, the topics were distinct so that subjects could indicate their familiarity with each text without ambiguity.

Phase 3: Topical Knowledge Advisors Recruitment

After the topics were determined, I recruited advisors from the related content areas to compose vocabulary lists of 10 words for each topic area. Two Taiwanese advisors from each content area were recruited for this study, for a total of eight advisors. The advisors who assisted me with the topics of law and engineering had obtained either masters or doctoral degrees in their respective fields from a prestigious university in Taiwan¹ and are currently working in their fields. The business topical knowledge advisors obtained their degrees in the U.S and are also working in their industries. The advisors in language teaching are associate professors in the English department of Tamkang University in Taiwan. These advisors have much experience within their disciplines. Their background information is provided in Table 3.

Table 3. Information About Content Advisors

Content Area	Educational Background	Years in Current Position
Law	1) Master's from National Taiwan, University's College of Law	4
	2) Master's from National Taiwan, University's College of Law	5
Language Teaching	1) New York University, Ph.D. in TESOL	12
	2) Tamkang University, Ph.D. in TESOL	7
Engineering	1) National Taiwan University, Master of Mechanical Engineering	4
	2) National Taiwan University, Ph.D. in Electronic Engineering	2
Business	1) The University of Iowa, Master of Business Administration	4
	2) New York University, Master of Economics	6

¹ Even in Taiwan, the textbooks used are originally written in English.

Phase 4: Vocabulary Generation

In this phase, I explained the purpose of the study and the requirements for potential participants in detail to each topical knowledge advisor who agreed to help. After reading the TOEFL practice test that was specific to their content area, the advisors were asked to write approximately 15 vocabulary words with definitions that they regarded as appropriate for someone familiar with that content area. The vocabulary was not limited to the exact words that were printed in the reading test text; other vocabulary words that were considered to be important were added to the list. For example, the vocabulary word “pedagogy” was not in the reading text for “reflection in teaching,” but it was added to the vocabulary assessment list by topical knowledge advisors.

Then, the two lists from each topical knowledge advisor in the same content area were reviewed and compared. I selected those vocabulary words that appeared on both advisors’ lists. To resolve any disagreements or confusion, I returned to the topical knowledge advisors for further consultation. Finally, 10 vocabulary words with advisors’ agreement in each area were generated for each category (see Appendix F). I submitted the final version of each vocabulary list to the topical knowledge advisors for their professional inspection and final opinions.

Phase 5: Vocabulary Scoring

The scoring system for the topical knowledge vocabulary assessment follows. The correct definitions provided by topical knowledge advisors were used as the criteria for scoring. As in previous studies by Huang (2010) and Valencia et al. (1991), participants scored one point for each vocabulary item correctly answered and zero points for definitions that were not related to the given vocabulary word in the specific content area. Those words answered partially were awarded a 0.5 score. The scores in each content area for every participant were summed as indicators of their topic familiarity.

The initial scoring was based on topical advisors' lists of definitions. After this initial evaluation, each topical knowledge advisor was given two randomly selected vocabulary tests to double check the scoring results to improve reliability.

Think-Aloud Protocols

Reading researchers have advocated the importance of reading processes because they help to clarify the reasons for the responses that readers make. However, the thoughts of readers are internal and are usually hidden from outside observers. Thus, oral reports from readers are necessary to gain insight into what readers are doing and thinking while completing a reading task. For this reason, verbal reports or think-aloud protocols have been used widely among researchers both in academic reading situations (Anderson, 1991; Bang & Zhao, 2007; Block, 1995) and in test-taking situations (Anderson, 1991; Cohen & Upton, 2007; Nevo, 1989; Tian, 2000).

By asking readers to reveal or explain the processes that they adopt in reading, the think-aloud protocol analysis has been a way of determining the cognitive strategies used by different readers during problem-solving tasks. Given that a multiple-choice reading comprehension test is a form of problem-solving activity, verbal reports can provide a better understanding of reading processes when they are used in conjunction with test scores (Green, 1998).

According to Ericsson and Simon (1984), there are two ways to conduct think-aloud protocols with problem-solving activities. The introspection protocol requires readers' thoughts to be reported item by item, but has the possibility to intrude and therefore distort reading processes. The retrospection think-aloud protocol, which requires readers to report their thoughts after the reading task is completed, was adopted in this study. The retrospective think-aloud protocol has the advantage of keeping the process and task intact but has the disadvantage that readers might forget what they were doing or thinking during their reading processes. Bowles (2010) suggested that this disadvantage can be "minimized if there is only a short delay between task performance

and verbalization” (p. 14). Given that the purpose of this study was to simulate the actual TOEFL iBT test without being intrusive, participants were asked to think-aloud immediately after time was up.

The use of think-aloud protocols is not without criticism. For example, readers might report only good strategies that they assume successful readers would use (Block, 1992). The strongest criticism is that there is no mechanism to confirm that the processes being reported are the actual processes being used (Bowles, 2010). Therefore, improvements have been made to ascertain the reliability and validity of the verbal report. For example, researchers may ask probing questions while obtaining the verbal report, as was done in this study. Participants may be required to provide reasons for selecting a certain option as well as their rationales for discarding other options. As a result, even with the criticisms noted, verbal reports have been widely defended for their advantages, given that strategies are mental, deliberate, cognitive actions that can only be accessed through conscious reporting (Anderson, 1991). For this reason, verbal reports or think-aloud protocols have been used extensively to gain insight into the cognitive processes that learners use while reading in their L2.

When a think-aloud protocol is used in a study of L2 learners, one consideration is whether the think-aloud protocol should be conducted in the L1 or the L2. Block’s (1986) pilot study required two ESL students to think aloud in their L2. Block reported that students responded with relative ease and their responses revealed information about their problems and strengths as readers. However, some investigators have found that more information was recalled in the student’s L1 because insufficient L2 proficiency might have hindered or limited their expression (Lee, 1986). Thus, when participants’ language proficiency differs, many studies allow students to choose their language use while participating in think-aloud activities (Abbott, 2006; Bowles, 2010; Cohen & Upton, 2007; Phakiti, 2003). To ascertain that participants are able to provide their

thoughts as completely as possible, it is advisable to allow both L1 and L2 during the think-aloud sessions.

Interviews

After administering a TOEFL iBT practice reading test, the researcher conducted a semi-structured interview to clarify ambiguous information about each participant's retrospective think-aloud protocol (see Appendix G). The semi-structured interview questions had the advantage of not containing predetermined, structured choices. Rather, the questions allowed individuals to lead the discussion, and the sequence of the interview questions could be changed to suit participants. Thus, the semi-structured interviews provided more flexibility.

First, participants' familiarity with the specific topic in the iBT reading practice tests was triangulated on a self-reporting scale from 1 to 10. Then the interviewer asked participants questions regarding their reading patterns and the order in which they answered the multiple-choice questions. Third, participants were asked to point out the easiest and most difficult questions with explanations of their responses to both familiar and unfamiliar texts. Fourth, participants' experiences in approaching familiar or unfamiliar texts were explored. The last interview question focused on how the reading processes compared if participants took a multiple-choice reading test and read in a non-testing situation.

Data Collection Procedure

A brief introduction to the purpose of this research and its procedures was emailed to students from both Taiwan and China who were enrolled at the University of Iowa. The content included the purposes of the study, the procedure, and the amount of time to complete participation. The value of students' participation and their cooperation was emphasized. Participants did not receive financial compensation, but received the scores on their practice tests and some feedback on how to improve their reading. When 30 students had agreed participate, data collection began.

Informational Session: Introduction, Retrospective
Think-Aloud Practice, and Topical Knowledge
Vocabulary Assessment

Those students willing to be participants in this study were gathered in a conference room to receive further instructions. Directions were given mainly in their native language, Mandarin Chinese. Information on the purpose, procedure, and duration of the study was explained to them orally to check if there was any misunderstanding. Students signed an informed consent form if they agreed to participate (see Appendix H).

Participants were introduced to the retrospective think-aloud protocol, which asked participants to report their thoughts after the reading task was completed. A similar audio-recorded demonstration of how a think-aloud protocol works, taken from the pilot study, was played for these participants. Next, participants had the chance to practice a think-aloud protocol session using an English practice text with a multiple-choice items that the researcher had prepared. Considering the unfamiliarity of think-aloud protocols, the purpose of this introduction was to help participants to think aloud their behavior when reading. Some hints were given to participants as they began the think-aloud protocol, such as, “Please tell me what you are thinking when you are reading this passage.” If the participants stopped or hesitated when describing their thoughts, some probing questions were asked, such as, “Please tell me more,” “What do you mean by...?,” and “Is there anything else you would like to share?”

The practice think-aloud encouraged participants to speak more without cuing them regarding what they should say. The think-aloud protocol could be completed in Chinese, English, or a combination of the two languages.

After the introduction and practice with the think-aloud protocol, if there were no further questions about the think-aloud process, participants were administered the topical knowledge vocabulary test. Finally, participants provided times during which they were available to meet individually with the researcher.

Test-Taking Session: Practice Test and Retrospective

Thinking-Aloud

Although the TOEFL iBT reading test is computer-based, the study conducted by Choi, Kim, and Boo (2003) supported the comparability between computer-based language tests and the paper-based language test versions. Thus, it was acceptable to adopt paper-based printed reading texts in this study, which included the same content as the computer-based TOEFL iBT practice test.

Participants met with the researcher individually in an empty classroom or in a conference room according to their availability. Participants were required to take two reading tests, one on a topic with which they were familiar and the other on a topic with which they were unfamiliar. Participants' familiarity with the four selected content areas was judged by their vocabulary assessment results. After conducting the participants' topical knowledge vocabulary assessment during the informational session, the researcher scored the participants' responses using the correct answers provided by the content advisors. Scores were summed as an indicator of every participant's topical knowledge about the four content areas (law, business, language teaching, and engineering).

After the results of the topical knowledge measurement were tabulated, it was possible to identify the iBT practice reading test for each participant that best fit the conditions of this study. Based on participants' performance on the vocabulary topical knowledge assessment, two specific content areas, including one text that was most familiar and one that was least familiar, were selected for each participant. Consequently, participants were tested on their own unique combinations of two reading passages corresponding to this experimental condition. Four participants were discontinued during this phase because they scored similarly on all four content areas, which did not fit with this research context.

When the testing session began, participants were given previously prepared and unique folders that included their particular sets of iBT reading practice tests. The order

of the two tests was randomly selected. When participants completed the iBT practice tests, the researcher sat next to them to observe and take notes on their physical behaviors. The time was limited to 20 minutes for each test, as allowed by the ETS, because the purpose of this task was to measure participants' reading skills in a normal standardized exam situation.

Participants were told to complete as many of the questions as they could within 20 minutes. Participants were asked to produce the think-aloud processes after they finished the first reading test. The think-aloud protocol was recorded. A translation into English was made by the researcher if the think-aloud protocol was given in Chinese.

There was a break of 10 minutes between the think-aloud protocol and the next test, if requested by the participants. This procedure was repeated with the second test.

After the second think-aloud protocol, a follow-up semi-structured interview was conducted to explore participants' patterns of answering questions in the reading assessment. If inconsistencies existed between participants' responses and the researcher's observations, further investigation was conducted to confirm participants' responses. The interviews were audio recorded and transcribed later for further analysis. The data collection procedure in this study is illustrated in Figure 2.

Topical Knowledge and Reading Comprehension

Assessment Results

To confirm that all participants fit the conditions of this study, participants' scores on the topical knowledge assessment and the two iBT practice reading comprehension tests were calculated. The process of elimination related to each assessment is described in detail below.

Each participant's topical knowledge vocabulary assessment was calculated. In this study, familiar readers were selected as participants who scored above 75% and unfamiliar readers were those who scored below 25% on any two content areas on the topical knowledge assessments.

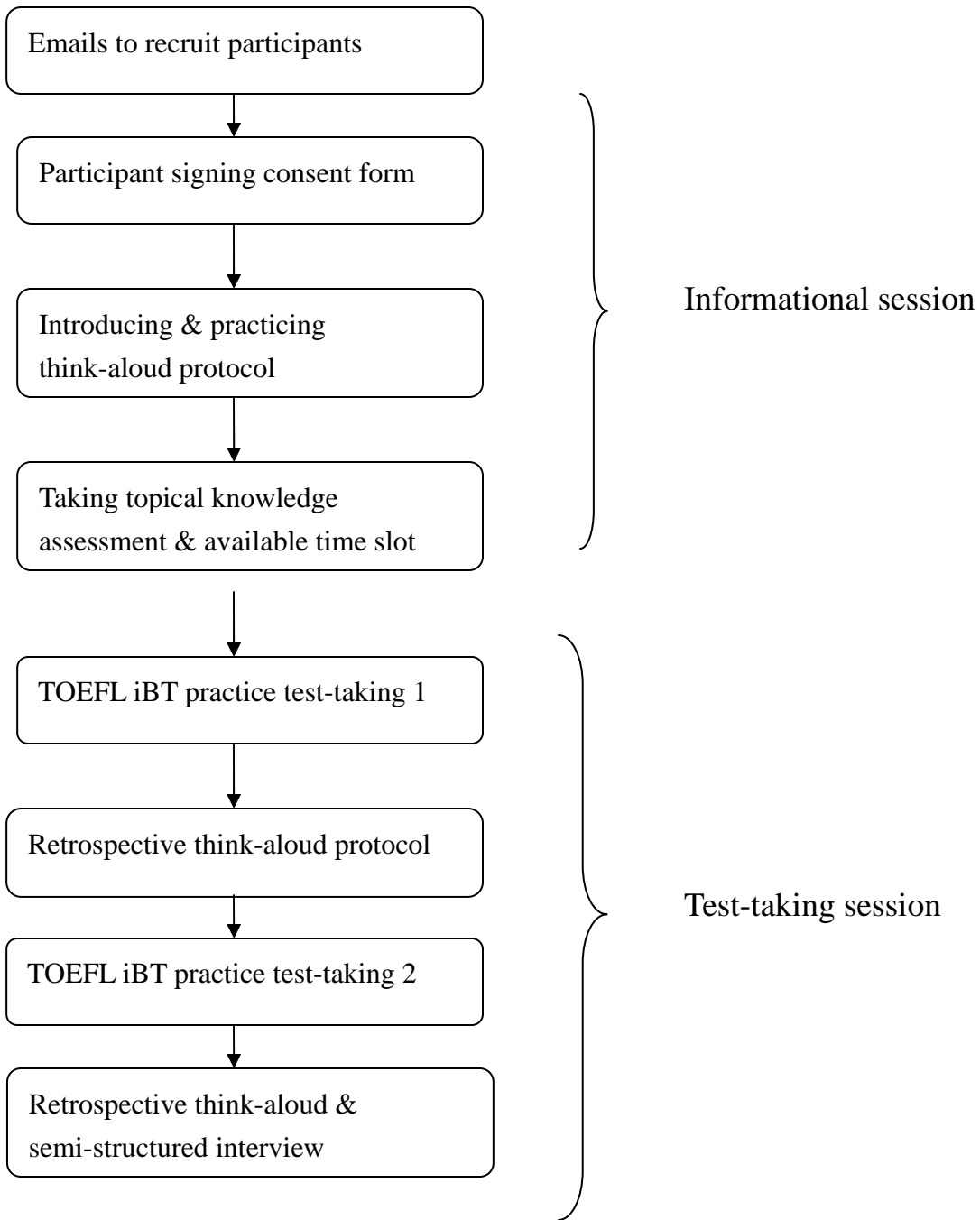


Figure 2. Data collection procedures

Four participants who scored similarly on all of the content areas were eliminated because they did not fit the conditions of this study. In the second participant elimination phase, the high and low level groups of participants were compared by their combined test scores on the two iBT practice tests, which totaled 20 points. Participants who obtained 70% or above (14 points or higher) were grouped as high-level readers, and those who scored below 30% (nine points or lower) were grouped as low-level readers. Those who scored in the middle range (seven to 13 points) of the iBT practice tests were eliminated. Finally, scores from the two combined iBT tests showed that 12 participants were grouped as high-level readers and eight were grouped as low-level readers for a total of 20 participants.

In summary, among the 30 participants, 10 were removed from the data analysis because they did not fit the conditions of this study. Four students were eliminated from further participation because they scored similarly on all of the content areas, and six students were eliminated because as they scored neither high nor low on their reading comprehension tests. The total number of participants who completed this study was 20. The topical knowledge assessment and reading comprehension test results for the 20 participants are presented in Tables 4 and 5.

Data Analysis

Based on the implemented design features, two stages of analysis, qualitative and quantitative, were undertaken in this study. The mixed method data analysis provides more comprehensive evidence than either qualitative or quantitative analysis alone. Using both types of data broadens the research analysis and therefore strengthens the results. The results can lead to more powerful generalization without being restricted to the types of data collection and analysis typically associated with qualitative or quantitative research (Creswell & Plano Clark, 2011).

Table 4. Results of Topical Knowledge Vocabulary Assessment

Participant's ID	Familiar Topic/ Score	Unfamiliar Topic/ Score
3	Business / 10	Language Teaching / 2
4	Language Teaching / 8	Law/ 2.5
5	Business / 7.5	Law/ 2
7	Engineering / 8	Law/ 2.5
8	Language Teaching / 9	Engineering / 1
9	Engineering / 10	Law/ 2
10	Engineering / 8	Language Teaching / 1
12	Language Teaching / 8.5	Engineering / 1.5
13	Business / 7.5	Law/ 1
14	Business / 9	Law/ 2
16	Law/ 7.5	Engineering/ 2.5
17	Language Teaching / 8.5	Law/ 2
18	Engineering / 7.5	Business / 1.5
21	Law/ 10	Engineering / 1
23	Law/ 8	Engineering / 2
25	Engineering / 8	Language Teaching / 2
26	Business / 10	Language Teaching / 2.5
27	Language Teaching / 9	Law/ 2
28	Language Teaching / 10	Business / 1
29	Language Teaching / 9.5	Business / 1.5

Table 5. Results of Combined iBT Reading Comprehension Tests

Participants' ID/ Level	Familiar + Unfamiliar score	Total score
3/ High	10 +7	17
4/ High	8 +9	17
5/ Low	3+ 5	8
7/ High	9 +9	18
8/ High	7 +8	15
9/ High	6+ 10	16
10 / Low	4+3	7
12/ High	8 +8	16
13/ Low	3 +5	8
14/ Low	3+ 4	7
16/ Low	3+ 3	6
17/ High	9+ 7	16
18/ Low	3 +2	5
21/ High	10+8	18
23/ Low	4 +2	6
25/ Low	2+3	5
26/ High	8 +9	17
27/ High	9 +9	18
28/ High	9 +10	19
29/ High	9+ 7	16

This study combined both qualitative and quantitative research and data analysis, but emphasized one data analysis more than the other. That is, quantitative data were used to elaborate on the qualitative data results, which were the focus of the study. This study combined qualitative and quantitative analyses and interpreted the results based on the integration of data analysis.

In this study, I analyzed the data in the following phases according to the specific mixed methods described above. In phase one, I calculated participants' scores on their topical knowledge vocabulary assessment and reading comprehension tests. Next, I collected and analyzed the qualitative data from those 20 remaining participants'

think-aloud protocols and interview responses. In phase three, I submitted the qualitative information obtained from phase two to the statistical analysis. Finally, I combined the qualitative and quantitative results and interpreted them together, with a closer examination of four selected participants. Qualitative data analysis and quantitative data analysis are discussed below in detail.

Qualitative Data Analysis

All verbal reports provided by participants in the think-aloud protocol and the semi-structured interview were recorded and transcribed by the researcher word for word. To develop a clear coding system for participants' strategy use, each participant's responses to both familiar and unfamiliar texts were reviewed several times by the researcher. Each response was analyzed to determine if any part of it met the definition of a strategy, and a label was assigned to any strategy use noted in the response. In this way, the researcher developed categories of processing strategies as well as examples for those strategies that were not as self-explanatory.

When the researcher had compiled a list of all strategies, previous research was consulted to determine if the strategies used by participants in this study had been addressed in the literature and, if so, how they were classified. References were made to previous research related to strategy categories and terminology (Alsheikh, 2011; Cohen & Upton, 2007; Farr et al., 1990; Pritchard, 1990; Tian, 2000). For example, Alsheikh (2011) defined three broad categories of strategies: global reading strategies, problem solving strategies, and support reading strategies. Cohen and Upton (2007) listed strategies in three categories: reading strategies, test-management strategies, and test-wiseness strategies. Farr et al. (1990) categorized: overall strategies, reading strategies, and test-taking strategies. Pritchard (1990) categorized five groups: developing awareness, accepting ambiguity, establishing intrasentential ties, establishing intersentential ties, and using background knowledge. Tian (2000) used three categories: technical strategies, reasoning strategies, and self-adjustment strategies. With the

previous research in mind, in my study, similar strategies were categorized into similar broad categories with a total of six for further statistical analysis.

The list of strategies identified from the think-aloud protocols and interview responses was given to two reading advisors with extensive experience in that area to conduct validity checks. To be more specific, the reading advisors were asked to check whether any discrepancies existed between (a) the think-aloud protocols and interviews and (b) the strategies identified by the researcher. Both advisors agreed that 39 out of 41 strategies represented valid ways of categorizing the data. After further discussion, two strategies were eliminated due to overlap. Finally, a professor who specialized in reading helped to determine the categorization of strategies.

In summary, the qualitative analysis of this study included (a) the verbatim transcript of all participants' think-aloud protocols and interviews, (b) the creation of categories of related strategies based on the think-aloud responses provided by all participants with regard to familiar and unfamiliar texts, and (c) the patterns of text-reading and multiple-choice question-answering sequences for each participant from their interviews.

Quantitative Data Analysis

Although this study's analysis was primarily qualitative, the researcher conducted some quantitative analyses to help consolidate the findings. The score for each participant's topical knowledge vocabulary assessment and two iBT practice reading comprehension tests was calculated. The frequency of each category reported in the think-aloud and interview responses was also calculated.

An analysis of variance (ANOVA) with repeated measures was conducted to determine if there were significant differences among group mean scores based on the frequency of strategy use and topic familiarity. To convert the positively skewed distribution to an approximate normal distribution, a required assumption for ANOVA, the square root of the frequency of each strategy was adopted.

The second ANOVA employed a split-plot statistical design to examine whether there were significant differences among group mean scores based on participants' strategy use and their reading scores. An analysis of variance was conducted with frequency of strategy use as the dependent variable and level of scores as the independent variable. Specifically, a 2 x 2 split-plot factorial design was employed, with reading scores (high and low performance) as the between-subject effect and frequency of strategy use as the within-subject effect.

Pilot Study

To identify possible problems with data collection procedures, a pilot study was conducted to test the research instruments and procedures during Winter 2011 with six Taiwanese students. Among them, two were graduate students who had experience with taking the TOEFL iBT test; the others had no experience or had taken the TOEFL CBT. Participants from the pilot study were chosen based on their similarity to the student sample from which the official participants would be recruited. The pilot study followed the same study procedure and research instruments.

For the pilot study, six Taiwanese students (two undergraduate students and four graduate students) completed the topical knowledge test, the two TOEFL iBT practice tests with different levels of familiarity, the retrospective think-aloud, and the follow-up interview. The researcher conducted the procedures with these six students to learn if changes were warranted in either the practice tests or the interview questions. First, in terms of the TOEFL practice test, participants were allowed 20 minutes to complete each text and to answer 10 multiple-choice questions. All participants found this length of time to be reasonable, especially for familiar texts. For instance, they indicated that the familiar text was easier and less fatigue-inducing due to their background knowledge. Second, because the TOEFL practice test resembled the on-line test with questions in the middle next to the paragraph, students who had not taken the iBT test reported having difficulty visually locating the questions. In response to this concern, the researcher

identified the questions for each participant before administering the test. Third, participants seemed to have no problems with refreshing their memories and performing retrospective think-aloud protocols. All participants preferred using Chinese. The think-aloud protocol took approximately 10 minutes for each text and 10 minutes for the students' interview responses. The total amount of time for completing this study was about 90 minutes for each participant.

To conclude, the results of the pilot study provided insights into the appropriateness of the test duration in that the results supported the utilization of semi-structured interviews associated with think-aloud protocols. As the results showed, the instruments and texts were sensitive enough to provide meaningful and reliable results.

The next chapter reports and discusses the results of this study.

CHAPTER 4

RESULTS

The main purpose of this chapter is to present the strategy use of Mandarin Chinese-speaking students when confronted with familiar versus unfamiliar topics in a multiple-choice format reading test based on the TOEFL iBT practice reading comprehension exam. The study examined test-takers' reading patterns and strategies through their think-aloud protocols and interviews. These strategy use reports were further examined with respect to participants' reading test scores.

Research Questions

The research is designed to address the following questions:

1. For academic English as a second language multiple-choice reading comprehension tests, what comprehension strategies do Mandarin Chinese readers use when reading about familiar topics?
2. What comprehension strategies do native speaking Mandarin Chinese readers use when reading about unfamiliar topics?
3. How do the strategies compare when reading about familiar versus unfamiliar topics?
4. Do strategies differ between students with high and low scores on the reading test?

The following section presents the results from the study with graduate students at the University of Iowa who responded to multiple-choice reading comprehension questions on a simulated high-stakes assessment. The findings of this study are presented in the context of the research questions.

Strategies and Topic Familiarity

Research Questions 1 and 2 discussed the test-taking strategies participants adopted when reading about familiar and unfamiliar topics in TOEFL iBT practice reading comprehension tests.

Distribution of Strategy Use

Participants' think-aloud protocols and interviews were used for determining strategy categories when they were reading familiar and unfamiliar topics in TOEFL iBT practice texts. Overall from the researcher's inspection of the data divided by topic familiarity, participants in this study used similar strategies across topics, regardless of familiarity. In other words, no differences were found in test-taking strategies for familiar versus unfamiliar topics, and the reading process used in comprehending across texts remained similar. For this reason, the results of Research Questions 1 and 2 are presented together. Further comparisons are detailed later in the section on Research Question 3.

Six broad categories of strategies that emerged from the think-aloud protocols were identified in this study: (a) general approaches to reading the passages, (b) identification of important information by the discourse structure of the passages, (c) vocabulary/sentence-in-context approaches, (d) multiple-choice test-management strategies, (e) test-wiseness, and (f) background knowledge. Each category of strategy is introduced in the following sections with explanations and examples.

Strategy Category 1

The first category, general approaches to reading the passages, referred to the overall sequences that participants employed when completing the TOEFL iBT practice comprehension tests, such as reading the passages first or the multiple-choice questions first. Also, this category considered how much, or what specific part of the passage, was read before or after attempting to answer the questions. Given that every participant tried to make the most efficient use of time in order to receive the highest score, individual differences during test-taking were found, and these differences were related to various sequences of how each participant completed the iBT reading task. In addition, participants may have attempted the reading test using their individual preferences for reading or test-taking, regardless of the topic they were reading. Generally speaking, strategies in this category were deliberate and goal-oriented, as participants planned in

their minds how to complete the iBT reading task. Strategies in this category also demonstrated participants' monitoring of test-taking processes and their efforts to remediate problems encountered by adjusting comprehension and adopting fix-it strategies. Nine strategies were found in the think-aloud protocols and interview responses in this first category. In the following lists, examples are provided when the strategy statement is less transparent or not self-explanatory. Also, an English translation is provided if the think-aloud or interview responses were given in Chinese. Otherwise, a direct quote was taken from the participant.

1. The entire passage is read first. Then each question is answered by going back to the related paragraph and looking for clues.

Ex: 我會先把全文看一遍，然後開始答第一題，假如無法回答再回去文中尋找相關段落，在繼續下一題。

I read through the entire article first. After reading all the text, I started to answer questions from the first. I would go back to reread the related paragraph which helps to locate information if I can't answer immediately. I answered the questions sequentially. (Participant 28, interview)

2. The questions that belong to the same passage are read together. Then the related passages are read with a search for the answers.

Ex: 我會先把同一段落的題目(或是題目群)先看完，再去讀那一段，然後把這幾題做完。

I read all the questions related to the same paragraph first. For example, questions 1 and 2 both belong to paragraph 1. Then I tried to find the clues from the relevant paragraph and go on to the next questions. (Participant 3, think-aloud protocol)

3. Only one question at a time is read and answered by reading the related passages and then searching for answers. Repeat.

Ex: 我先看題目，而且一次只看一題，然後回相關文中找答案，答完之後再答下一題。

I read the first question and looked for clues in the related paragraph. Then I went to the next question (2, 3...) with the same procedure. (Participant 7, interview)

4. The participant reads a portion of the passage (that might contain a potential answer) carefully.

Ex: 雖然關鍵字上一段就出現，可是我知道這題答案在下一段，因為下一段才有指出為什麼作者要提出這些例子。

Although the key word already showed up in the first paragraph, I still went to the next paragraph, for it explained what wasn't mentioned in the first paragraph: the reason why the author raises the example, which should be the answer to this question. (Participant 12, think-aloud protocol)

5. The participant reads rapidly/ skims/ skips the passage if questions are not being asked.

Ex: Since I have already found this answer, and I know the next questions belonged to the next paragraph, I skimmed or even skipped the rest of the paragraph. (Participant 21, think-aloud protocol)

6. The participant looks for markers of meaning in the passage (e.g., quotes, bold text, people's names, numbers, or definitions).
7. The participant paraphrases and translates words, phrases, or sentences into Chinese.
8. The participant rereads certain paragraphs to clarify the idea.
9. The participant calculates in his/her mind how much time is left.

Ex: 時間已經不多了，我要趕快把剩下的題目做完。

I don't have too much time left. I need to hurry up with my reading and distribute the remaining time evenly in order to be able to finish the rest of the questions in time. (Participant 12, interview)

Strategy Category 2

The second strategy category, identification of important information by the discourse structure of the passage, revealed participants' understanding of English discourse structure, and they used this knowledge to analyze how the text was organized as a whole. Participants adopted these strategies by taking advantage of this hierarchical linguistic knowledge in English when comprehending an English text, utilizing grammatical, lexical, syntactic, and semantic rhetorical patterns in language. Four strategies were found in this category.

1. The participant looks for sentences that convey the main ideas.

Ex: 我會把段落的主題句跟結尾句仔細看一下，因為那幾句通常代表整段的最重要意義。

I would focus my attention to the first few sentences and the last few sentences in each paragraph, since those sentences indicate the main points of the whole paragraph. (Participant 21, interview)

2. The participant uses knowledge of the discourse genre of the passage /portion of the text (cause/effect, comparison/contrast, etc.).

Ex: The first word in this paragraph started with “however”, so I know the meaning should be different from the previous part and the author is going to bring out a new argument. (Participant 29, interview)

3. The participant uses knowledge of organization patterns and notes the different parts of the passage and how they interrelate.

Ex: 在我閱讀的時候，我很清楚文章的框架和段落之間的關係。譬如說這是開頭介紹，然後例子，再轉折到另一個概念等等.....

When I was reading, I was very aware of the framework and different parts of the passage. For example, this is the introduction, the examples, the transitions, and the conclusion, etc. (Participant 12, interview)

4. The participant uses knowledge of logical connectors to clarify content and passage organization.

Ex: 在我閱讀時，我會特別注意這種邏輯性的轉折詞，像是首先、此外、然而、最後結論...

While I was reading, I paid special attention to those connectors, such as: first of all, on the other hand, in conclusion, that is...etc. (Participant 4, think-aloud protocol)

Strategy Category 3

The third category that emerged from the think-aloud was vocabulary/sentence-in-context approaches. One type of iBT test question measures test-takers’ ability to understand the meaning of the selected words or sentences in the text as well as to insert a new sentence into a section of the text. The strategies in this category were adopted when participants made inferences about the unknown vocabulary words or sentences from the semantic or syntactic clues in context, and they tried to make inferences about the inserted sentences from their neighboring content. In this case, these strategies were local and text-based, which focused on isolated vocabulary, phrases, and sentences. Six strategies were found under this category:

1. The participant looks at sentences before or after it to verify the referent of a pronoun.
2. The participant infers/confirmes the meanings of new words through prefixes.

Ex: 我不認識 habitual 這個字，但我猜他跟是 habit 的形容詞，跟習慣有關 所以我去選項裡找，reoccurring 有 re 字根，代表也是重覆。

I didn't know the word "habitual." But I guess it is related to the word "habit." As I went through the four options, I found the word "reoccurring." The prefix "re" means again and repetition, so I selected this one. (Participant 26, think-aloud protocol)

3. The participant infers/confirmes the meanings of new words through semantic clues from the context.

Ex: 這題問我 emit 這個字的意思，而我不知道它的意思是什麼，所以我單字所在的句子看一次，emit 二氧化碳和有毒物質，那意思應該是釋放。

This question asked me the meaning of the word "emit." I didn't know this word, but from the sentence, it said "emit carbon dioxide and air pollutant..." So this word should mean "release." (Participant 12, think-aloud protocol)

4. The participant infers/confirmes the meanings of new words through syntactic clues from the context.

Ex: 這個句子說 good or even...所以後面一定是接比 good 還要好

The sentence has its structure as "good or even..." So I know that I should select the option which is very positive or better. (Participant 27, think-aloud protocol)

5. The participant infers/confirmes the meanings of the highlighted sentences through semantic clues from the context.
6. The participant infers/confirmes the meanings of new inserted sentences through meanings within the context.

Ex: 在第一個框框它有說“拿橘子來當例子”，所以是剛開始提到橘子，插入句不可能放這裡。後面繼續開始比較佛羅里達跟亞歷桑納州的氣候，來說明為什麼這兩地方適合橘子。之後再說拿橘子當例子，我覺得才是應該插入的地方。

Near blank 1, the sentence mentioned "take oranges for example..."; it was just the beginning of the paragraph, and was not ready to introduce something about the orange. And the following sentences compared the weather conditions in Florida and Arizona, which explained if these two places were suitable for orange

planting. So that was related to the sentence needed to be inserted. According to their relevance, I inserted the sentence here. (Participant 3, think-aloud protocol)

Strategy Category 4

The fourth category, multiple-choice test-management strategies, referred to how participants understood and answered the comprehension questions using clues given by either the questions or the options. As multiple-choice testing is a unique format with several answer options, participants sometimes caught and inferred the main points of the passage through the questions or the list of options. Eleven strategies were found to process information from this category. Two subcategories emerged: strategies by questions (1-6) and strategies by options (7-11).

Questions

1. The participant makes a mental note of the key points of the question and searches for the answer in text accordingly.

Ex: 這題題目重點是要問"tort cases"，所以我就去文章找有關的地方。

This question asked about "the cases of tort". So I highlighted the key word in my mind and looked for where the related text was located. (Participant 21, think-aloud protocol)

2. The participant matches a key word in the question to the text.

Ex: 這題題目有老師 concern什麼東西，所以我就想到去對應文中有提到worry這個字。

After reading the questions, I realized that the key words "concern" in the text and the word "worry" in the question corresponded with each other. (Participant 29, think-aloud protocol)

3. The participant extracts key meaning of text through questions.

Ex: This question asks why do defendants prefer being convicted in civil cases than in criminal cases? So I realized that the word "incarceration" is probably the key word of this paragraph. (Participant 21, think-aloud protocol)

4. The participant rereads or paraphrases the question for clarification.
5. The participant reads the question and jumps immediately to where the related text is, either before or while considering options.

6. The participant finds the key word/point/punctuation mark in the inserted sentence and connects it to the context.

Ex: Well...after reading the whole sentence, I know it's talking about this kind of orange, you need to pay this stuff...it's very detailed. Basically it's about the details relevant to orange. So you need to go back and find the very beginning when it is talking about the details of orange and connect with the context. (Participant 18, think-aloud protocol)

Options

7. The participant infers text meaning by considering the options first, then rereads the related text portion.
8. The participant extracts key meaning of text through list of options.
9. The participant selects options based on the paragraph/passage's overall meaning.

Ex: 因為我知道這段的主題都在講教師反思的益處，所以很輕鬆選了這個有關的答案。

I knew this whole passage talked about the benefit of teacher's reflection, so I selected this option because other options were irrelevant to this issue. (Participant 29, think-aloud protocol)

10. The participant selects the option because it appears to have a word or phrase from the passage in it – possibly a key word.
11. The participant paraphrases the options with text.

Ex: 這文章說到，老師很少有機會尤客觀的角度去檢視他們自己的教學”，然後我把文章意思換句話說，就是第二個選項：仔細檢視在教室裡所發生的一切...我想答案就是這個。

The text mentioned that “teachers rarely have the time or opportunities to view their own or the teaching of others in an objective manner,” so I began to paraphrase the text with the four options and to check if they corresponded to the text. For example, in the second choice, “examine thoughtfully the possible causes of events in their classrooms”...I think this should be the right one. (Participant 29, think-aloud protocol)

Strategy Category 5

The fifth category, test-wiseness, referred to participants' ability to utilize the characteristics and formats of the test to receive a high score without understanding or comprehending the passage. These strategies helped participants to select the correct

answer in the multiple-choice tests without knowing the content or using the language skills that should be tested. Five test-wiseness strategies emerged from this category.

1. The participant uses the process of elimination to achieve an answer (i.e., selecting an option even though it is not understood, out of a vague sense that the other options couldn't be correct).
2. The participant eliminates option(s) as contradictory to paragraph/overall passage meaning.

Ex: 這個選項絕對不是答案，因為他提到風力發電的負面影響，但是整篇文章就是在講風力發電的優點，所以一定不對。

This option could be eliminated first because it talked about disadvantage of wind power, which obviously contradicted the overall meaning of this paragraph. (Participant 7, think-aloud protocol)

3. The participant eliminates options as not mentioned in designated paragraph.

Ex: The option D could be eliminated first because it talked about the technology, an issue that did not appear in this paragraph. (Participant 17, think-aloud protocol)

4. The participant selects the option because of evenly distributing answers.

Ex: I remembered that the four answers will occur with equal frequency. So I chose A because it was selected less than other answers. (Participant 5, interview response)

5. The participant selects the longest answer.

Strategy Category 6

The sixth and the final category referred to the use of background knowledge about the content in the passage as the basis of comprehension and selection of an answer. Use of background knowledge has been recognized as an important factor in many studies related to comprehension processes. Block (1986) described this strategy as “use of general knowledge and associations.” In this study, the strategies in this category referred to the use of background knowledge of the content area discussed in the passage.

1. The participant considers prior knowledge before reading the text.

Ex: 這篇文章在講供需平衡，以前經濟課好像有學過一點。

This article talks about law of demand, which seems to be familiar to me as I took a class in economics before. (Participant 7, interview response)

2. The participant selects or eliminates options through background knowledge as an educated guess.
3. The participant predicts or produces his/her own answer (through background knowledge) after reading the portion of the text referred to by the question.
4. The participant predicts or produces his/ her own answer (through background knowledge) after reading the question and then looks at the options without referring to text.

Ex: 這題題目問說：只有在什麼條件之下，人們會去追尋奢侈要求？我就預期答案，就是當所有基本需求都滿足後。

This question asked when people only speak in terms of wants. I can predict the answer, that is, when people have already fulfilled their basic necessities. (Participant 23, think-aloud protocol)

In summary, when all the think-aloud protocols were coded, a total of 39 strategies were identified in the six broad categories: nine general approaches to reading the passages, four identification of important information by the discourse structure of the passages, six vocabulary/sentence-in-context approaches, 11 multiple-choice test-management strategies, five test-wiseness, and four background knowledge-related strategies. A list of all strategies is presented in Table 6.

Relationship Between Strategy Use and Topic Familiarity

The third research question compared the test-taking strategies employed when participants read texts about familiar and unfamiliar topics. As noted in the previous section, the strategies used for the two texts appeared similar. The average scores on the two texts were further calculated to investigate the influence of topic familiarity. The mean score was 6.25 for the familiar texts and 6.15 for the unfamiliar tests, which substantiated the findings that participants performed similarly regardless of topic

Table 6. TOEFL iBT Six Broad Categories of Strategies

Categories of Strategies	Definition
A. General approaches to reading the passage	<ol style="list-style-type: none"> 1. The entire passage is read first. Then, each question is answered by going back to the related paragraph and looking for clues. 2. The questions belonging to the same passage are read together. Then the related passages are read with a search for the answers. 3. Only one question at a time is read and answered by reading the related passages with a search for answers. Repeat. 4. Reads a portion of the passage (might be potential answer) carefully. 5. Reads rapidly/ Skims/ Skips the passage if not being asked. 6. Looks for markers of meaning in the passage (e.g., quotes, bold text, people name, numbers, definitions, examples). 7. Paraphrases and translates words, phrases, or sentences. 8. Rereads certain paragraph to clarify the idea. 9. Calculates in mind how much time left.
B. Identification of important information by the discourse structure.	<ol style="list-style-type: none"> 1. Looks for sentences that convey the main ideas. 2. Uses knowledge of the discourse genre of the passage (cause/effect, compare/contrast, etc.). 3. Uses knowledge of the organization patterns by noting the different parts of the passage (introduction, examples, transitions, etc.) and how they interrelate. 4. Uses knowledge of the connectors (first of all, on the other hand, that is, etc.).
C. Vocabulary/ sentence-in-context approaches.	<ol style="list-style-type: none"> 1. Looks neighboring sentences to verify the referent of a pronoun. 2. Infers/confirms the meanings of new words through prefixes. 3. Infers/confirms the meanings of new words through semantic clues from context. 4. Infers/confirms the meanings of new words through syntactic clues from context. 5. Infers/confirms the meanings of the highlighted sentence through semantic clues from context (and paraphrase the options). 6. Infers/confirms the meanings of new inserted sentences through meanings of context.

Table 6 (continued)

D. Multiple-choice test-management strategies.	<ol style="list-style-type: none"> 1. Makes a mental note of the key points of question and searches for the answer in text accordingly. 2. Matches key words in the question to the text. 3. Extracts key meaning of the text by questions. 4. Rereads /paraphrases the question for clarification. 5. Reads the question and jumps immediately to where the related text is, either before or while considering options. 6. Finds the key word/point/ punctuation mark in the inserted sentence and connect to the context. 7. Infers text meaning by considering the options first and rereads the related text portion. 8. Extracts key meaning of text through list of options. 9. Selects options based on paragraph/passage's overall meaning. 10. Selects the option because it appears to have a word or phrase from the passage in it – possibly a key word. 11. Paraphrases options with the meaning of the text.
E. Test-wiseness skills.	<ol style="list-style-type: none"> 1. Uses the process of elimination to achieve an answer even though the option is not understood. 2. Eliminates option(s) as contradictory with paragraph/overall passage meaning. 3. Eliminates options as not mentioned in designated paragraph. 4. Selects the options because of evenly distributing answers. 5. Selects the longest answer.
F. Use of background knowledge.	<ol style="list-style-type: none"> 1. Considers prior knowledge before reading the text. 2. Selects/Eliminates options through background knowledge as an educated guess. 3. Predicts or produces own answer (through background knowledge) after reading the portion of the text referred to by the question. 4. Predicts or produces own answer (through background knowledge) after reading the question and then looks at the options (before returning to text).

familiarity, and different topics did not seem to have a major influence on reading comprehension. To more completely answer this research question, the researcher used quantitative analysis to take a closer look at the frequency of strategy use between familiar and unfamiliar topics

Six strategy categories used when reading familiar versus unfamiliar texts in this study were identified and then their corresponding means as well as standard deviations

were calculated (see Table 7). Figure 3 shows a graph based on the results from Table 7. In addition, the summary statistics (means and standard deviations) for all strategies between the two task treatments, familiar and unfamiliar, are also presented in Table 7.

Table 7. Means and Standard Deviations of Familiar and Unfamiliar Categories of Strategies

Type	Strategy Frequency		Mean	Std Dev
	Strategy	N		
FAMILIAR	1	20	2.27	0.28
	2	20	1.35	0.52
	3	20	1.60	0.25
	4	20	2.46	0.23
	5	20	1.58	0.16
	6	20	1.18	0.33
UNFAMILIAR	1	20	2.32	0.23
	2	20	1.40	0.40
	3	20	1.65	0.18
	4	20	2.48	0.25
	5	20	1.62	0.21
	6	20	1.01	0.46
TOTAL	1	20	2.30	0.25
	2	20	1.37	0.46
	3	20	1.63	0.21
	4	20	2.47	0.24
	5	20	1.60	0.18
	6	20	1.10	0.41

For the familiar texts, the means of individual strategy use ranged from a high of 2.46 (multiple-choice test strategies) to a low of 1.18 (background knowledge strategies). Similarly, for the unfamiliar texts, the means of individual strategy use ranged from a high of 2.48 (multiple-choice test strategies) to a low of 1.01 (background knowledge

strategies). In terms of total strategies across topic familiarity, the means of strategy use ranged from a high of 2.47 to a low of 1.10.

Looking at Figure 3, the similarity in strategies with familiar versus unfamiliar texts is apparent. This is shown through a more frequent use of multiple-choice strategies and general reading approaches, while background knowledge strategies and discourse structural strategies were less frequently used.

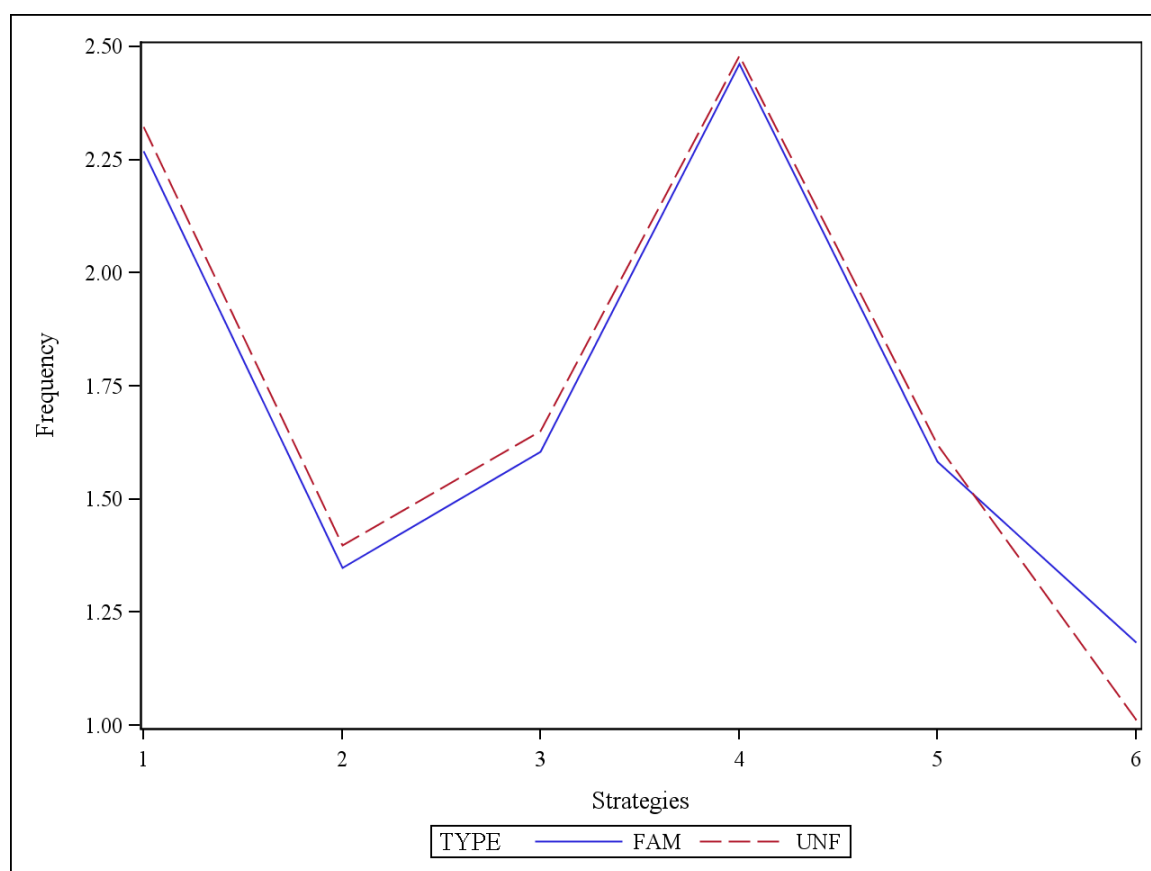


Figure 3. Six categories of strategies used in familiar versus unfamiliar texts

Analysis of variance (ANOVA) with repeated measures was carried out to determine if there were significant differences among group mean scores based on the frequency of strategy category use and topic familiarity. To convert the positively skewed

frequency distribution to an approximate normal distribution, a required assumption for ANOVA, the square root of the frequency of each strategy was adopted.

The results were summarized in Table 8. The 2 x 6 ANOVA analysis showed that the main effect of topic familiarity was not statistically different ($F=0.02$, $p =0.893$). That is, participants tended to use similar strategies when they were reading either familiar or unfamiliar texts. However, the main effect of strategy category was statistically significant ($F= 117.53$, $p < 0.0001$), which indicates difference in the employment of strategies across the six strategy categories.

Table 8. ANOVA Analysis of Topic Familiarity and Reading Strategy

Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	P Value
FAMILIARITY	1	19	0.02	0.8930
STRATEGY	5	95	117.53	<.0001
FAM*STR	5	95	0.80	0.5531

With the frequency of strategy use being statistically significant, a Tukey follow-up statistical analysis was conducted to examine the main effect for the six strategy categories (see Table 9). From the results, with the nominal 0.05 used throughout this study, there was no significant difference between strategies in categories 3 (vocabulary/sentence-in-context approaches) and 5 (test-wiseness skills) ($p = 0.7135$). However, this was not the case for other pairwise comparisons. These differences are illustrated in Figure 3.

More specifically, the mean score of strategy category 1 ($\bar{X} =2.30$) shown in Table 7 was significantly higher than strategy category 2 ($\bar{X} =1.37$), strategy category 3 ($\bar{X} =1.63$), strategy category 5 ($\bar{X} =1.60$), and strategy category 6 ($\bar{X} =1.10$) ($p < 0.05$). It indicated that the strategies in category 1 (general reading approaches) were used more

frequently than strategies in 2 (discourse structural strategies), 3 (vocabulary/sentence-in-context), 5 (test-wiseness), and 6 (background knowledge). The mean score of strategy category 2 ($\bar{X} = 1.37$) was statically significant lower than strategy category 1 ($\bar{X} = 2.30$), strategy category 3 ($\bar{X} = 1.63$), strategy category 4 ($\bar{X} = 2.47$), and strategy category 5 ($\bar{X} = 1.60$) ($p < 0.01$). It indicated that the strategies in category 2 (discourse structural strategies) were reported less frequently than strategies in 1 (general reading approaches), 3 (vocabulary/sentence-in-context), 4 (multiple-choice management), and 5 (test-wiseness). The mean score of strategy category 3 ($\bar{X} = 1.63$) was statically significant lower than strategy category 1 ($\bar{X} = 2.30$) and strategy category 4 ($\bar{X} = 2.47$), but higher than strategy category 2 ($\bar{X} = 1.37$) and strategy category 6 ($\bar{X} = 1.10$). It suggested that participants seemed to favor strategies in categories 1 (general reading approaches) and 4 (multiple-choice management) over category 3 (vocabulary/sentence-in-context), followed by strategies in categories 2 (discourse structural strategies) and 6 (background knowledge).

The mean score of strategy category 4 ($\bar{X} = 2.47$) was significantly higher than the rest of the categories of strategies in this study ($p < 0.01$), which suggested that strategies in category 4 (multiple-choice management) were the most frequently employed. The mean score of strategy category 5 ($\bar{X} = 1.60$) was higher than strategy category 2 ($\bar{X} = 1.37$) and strategy category 6 ($\bar{X} = 1.10$) ($p < 0.001$). The mean score of category strategy 6 ($\bar{X} = 1.10$) was significantly lower than all of the other categories ($p < 0.0001$), which indicated that the background knowledge strategies were least adopted by participants in this study.

Overall, participants in this study employed strategies in category 4 (multiple-choice test-management strategies) most frequently, followed by strategies in category 1 (general approaches to reading the passage), category 3 (vocabulary/sentence-in-context approaches), and category 5 (test-wiseness). Then, participants in this study employed strategies in category 2 (identification of important

information by the discourse structure of the passage) second from the bottom and employed strategies in category 6 (background knowledge) least frequently (see Table 8 and Figure 3).

Table 9. Results of the Tukey Follow-up Analysis

Effect	SUBSCALE	_SUBSCALE	Estimate	Standard Error	DF	t Value	Pr > t
STRATEGY	1	2	0.92	0.07	95	13.26	<.0001
STRATEGY	1	3	0.67	0.07	95	9.61	<.0001
STRATEGY	1	4	-0.17	0.07	95	-2.51	0.0139
STRATEGY	1	5	0.70	0.07	95	9.98	<.0001
STRATEGY	1	6	1.20	0.07	95	17.23	<.0001
STRATEGY	2	3	-0.26	0.07	95	-3.65	0.0004
STRATEGY	2	4	-1.10	0.07	95	-15.77	<.0001
STRATEGY	2	5	-0.23	0.07	95	-3.29	0.0014
STRATEGY	2	6	0.28	0.07	95	3.97	0.0001
STRATEGY	3	4	-0.84	0.07	95	-12.12	<.0001
STRATEGY	3	5	0.026	0.07	95	0.37	0.7135
STRATEGY	3	6	0.53	0.07	95	7.62	<.0001
STRATEGY	4	5	0.87	0.07	95	12.48	<.0001
STRATEGY	4	6	1.34	0.07	95	19.74	<.0001
STRATEGY	5	6	0.50	0.07	95	7.26	<.0001

Although the test-taking strategies used by each participant across texts with different topic familiarity were very similar, the unfamiliar text prompted a modest increase in test-taking strategies use in this study, but not in a statistically significant way.

In addition to the quantitative counts of strategy use from participants' think-aloud protocols, the qualitative interview responses also revealed that topic familiarity did not account for difference in strategy selection. Most of the participants noted that because the iBT reading section is designed to evaluate a test-taker's understanding of textual information instead of one's background knowledge, their top priority was to comprehend the text by utilizing their reading skills.

Ex: 畢竟托福考試是考閱讀理解能力，而不是你的背景知識，所以背景知識只是有幫助，但不會影響整體的閱讀方式。

Since the TOEFL reading section has its main purpose for testing ones' reading comprehension, not one's background knowledge about a specific topic or content. Therefore, I still focus on my comprehension and will use the same types of strategies to comprehend the text. (Participant 21, interview response)

The only difference related to topic familiarity retrieved from the interview responses was that participants tried to connect the familiar text to what they had learned previously. However, having the general background knowledge was not enough to understand an iBT text that contained specific information. For those test-takers who could not rely on their background knowledge about a topic, they could still obtain the correct response when comprehending the text successfully, as expected. Thus, background knowledge inevitably can help at some point in reading assessments, but these chances are rare. Participants all reported that their ultimate goal was to comprehend the text to obtain satisfactory test results.

Another topic familiarity issue revealed in the interviews was that the familiar text played an important role in relieving tension, especially in a high-stakes exam. This affective impact was mentioned by all the participants interviewed. That is, the background knowledge impacted test-takers' emotional factors more than their reading patterns. As expected, participants felt more confident when they had the opportunity to read passage that was related to what they had learned. Even when they could not understand every single word in the text, they believed they controlled their reading

processes more effectively when reading familiar texts. One of the participants described the experience:

Ex: 閱讀熟不熟悉主題對我來說，所使用的策略都是一樣的。唯一不同的就是考試感覺吧，閱讀熟悉主題會比較有自信，就算不了解每個字意，也可以用背景知識來聯想。

I used the same test-taking strategies across different topics. The only difference lies in the feeling when I am answering the questions. I may feel more confident when I read familiar articles, because I can think about it in a broader way. Even though I don't read through every single word, you have the general background knowledge. (Participant 29, interview response)

Strategies and Test Score

The last research question in this study compared the strategy selection among high and low scorers. Each participant's test scores of the familiar and unfamiliar text were summed as an indicator of their performance on the practice TOEFL iBT reading tasks.

Relationship Between Strategy Use and Test Score

Six categories of strategies used by 12 high and eight low scorers in this study were counted and then their corresponding means and standard deviations were calculated (see Table 10).

The second ANOVA employed a split-plot statistical design to examine whether there were significant differences among group mean scores based on participants' strategy use and their reading scores. An analysis of variance was conducted with frequency of strategy use as the dependent variable and level of scores as the independent variable. Specifically, a 2 x 2 split-plot factorial design was employed, with reading scores (high and low performance) as the between-subjects effect and frequency of strategy use as the within-subjects effect (see Table 11). In the 2 x 6 split-plot ANOVA of reading performance and frequency of strategy use, the interaction of the two main effects was significant ($F=11.14, p < .0001$). This indicates that there were significant differences in strategy patterns between the high and low scorers when performing the reading comprehension tests. Given that this research question compared the strategy

selection by high and low scorers, the follow-up test focused only on this aspect of the ANOVA analysis.

Table 10. Means and Standard Deviations of Strategies by High and Low Scorers

STRATEGY FREQUENCY				
PERFORMANCE	STRATEGY	N	Mean	Std Dev
High	1	12	2.95	0.11
	2	12	1.93	0.10
	3	12	2.11	0.10
	4	12	3.14	0.12
	5	12	1.80	0.17
	6	12	1.38	0.47
Low	1	8	2.41	0.16
	2	8	1.26	0.57
	3	8	1.69	0.12
	4	8	2.63	0.13
	5	8	2.08	0.11
	6	8	1.36	0.14

Table 11. Results of the 2 x 6 Split-plot ANOVA for Reading Scores and Strategy Use

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	P Value
PERFORMANCE	1	18	49.15	<.0001
STRATEGY	5	90	121.09	<.0001
PER*STR	5	90	11.14	<.0001

Follow-up simple effect tests (see Table 12) showed that there were significant differences in strategy use between high and low scorers for strategy category 1 ($F=23.97$, $p < .0001$), strategy category 2 ($F=37.05$, $p < .0001$), strategy category 3 ($F=15.46$, $p = 0.0002$), strategy category 4 ($F=21.69$, $p < .0001$), and strategy category 5 ($F= 6.66$, $p = 0.0115$). In terms of strategy category 6, no significant difference was found between high and low scorers ($F=0.03$, $p = 0.8554$). The high and low scorers in this study were different in their strategy selection in categories 1 (general reading approaches), 2 (discourse structural strategies), 3 (vocabulary/sentence-in-context), 4 (multiple-choice management), and 5 (test-wiseness), but not in category 6 (background knowledge).

Table 12. Results of Follow-up Simple Effects Tests

Tests of Effect Slices						
Effect	STRATEGY	Num DF	Den DF	F Value	P Value	
PER*STR	High	5	90	99.12	<.0001	
PER* STR	Low	5	90	44.11	<.0001	
PER* STR	1	1	90	23.97	<.0001	
PER* STR	2	1	90	37.05	<.0001	
PER* STR	3	1	90	15.46	0.0002	
PER* STR	4	1	90	21.69	<.0001	
PER* STR	5	1	90	6.66	0.0115	
PER* STR	6	1	90	0.03	0.8554	

Figure 4 presents a graph based on the results from Table 10. The mean score for high scorers in strategy category 1 ($\bar{X} = 2.95$) shown in Table 10 was significantly higher than for low scorers ($\bar{X} = 2.41$), which indicated that high scorers employed strategies in

category 1 (general reading approaches) more frequently than low scorers. The results were repeated in strategy category 2 ($\bar{X} = 1.93 > \bar{X} = 1.26$), strategy category 3 ($\bar{X} = 2.11 > \bar{X} = 1.69$), and strategy category 4 ($\bar{X} = 3.14 > \bar{X} = 2.63$). These differences suggested that high scorers employed a comparatively higher frequency use of strategies in category 2 (discourse structural strategies), 3 (vocabulary/sentence-in-context), and 4 (multiple-choice management).

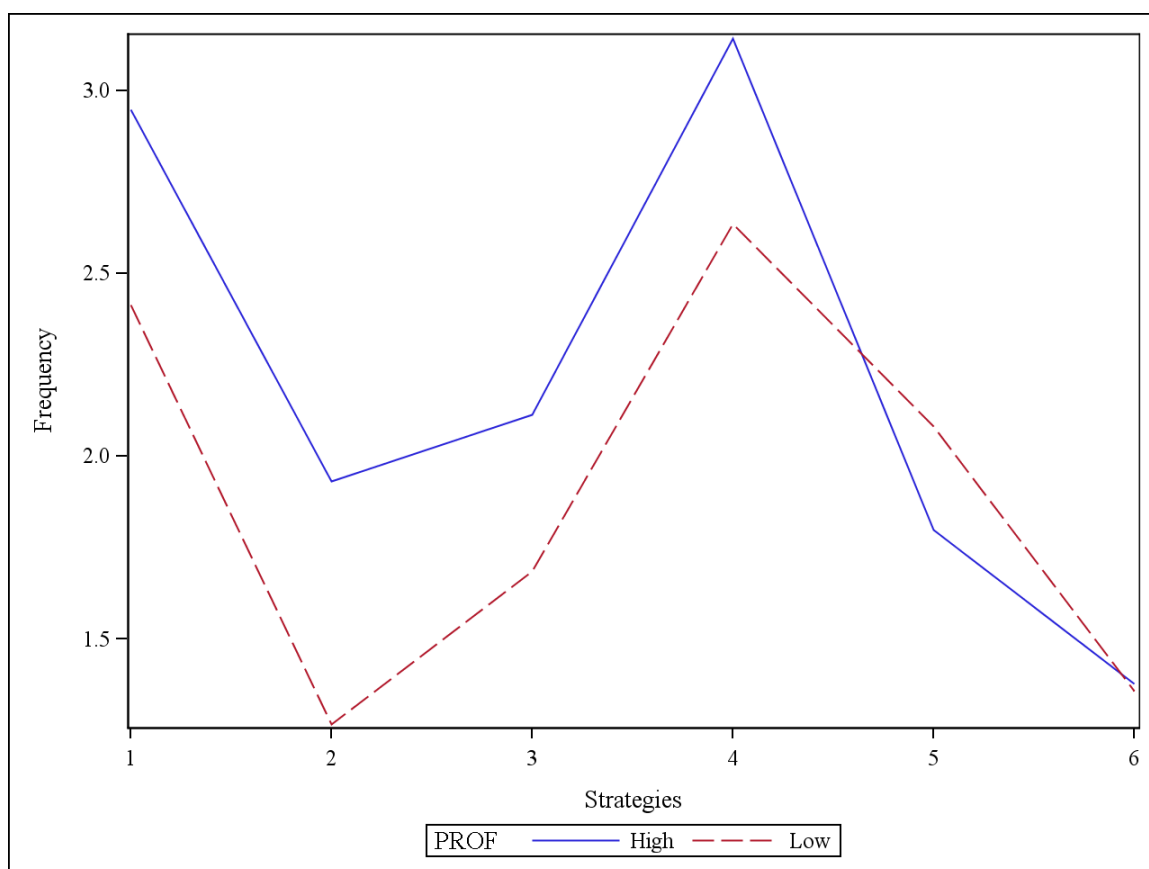


Figure 4. Six categories of strategies used by high and low scorers

However, the mean score for high scorers in strategy category 5 ($\bar{X} = 1.80$) was significantly lower than for low scorers ($\bar{X} = 2.08$), indicating that strategies in category 5 (test-wiseness) were used more often by low scorers. In terms of category of strategy

category 6 (background knowledge), the mean scores for high scorers ($\bar{X} = 1.38$) and low scorers ($\bar{X} = 1.36$) were not significantly different, which suggested that the background knowledge strategies were used a similar number of times by high and low scorers in this study.

Generally speaking, high scorers employed strategies significantly more frequently than low scorers in category 1 (general approaches to reading the passage), 2 (identification of important information by the discourse structure), 3 (vocabulary/sentence in context approaches), and 4 (multiple-choice test-management strategies). However, low scorers significantly adopted more strategies in category 5 (test-wiseness). Concerning the last category, high and low scorers seemed to use a similar number of strategies in category 6 (background knowledge).

Case Studies of High and Low Scorers

To investigate the individual nature of strategy use, four case studies are presented to illustrate the test-taking strategies that led to success or failure in this simulated iBT reading task. The data for the case studies came from participants whose test scores indicated very different test-taking processes but who scored similarly: two participants who scored high and two who scored low on both types of reading comprehension texts. Pseudonyms are used for each participant.

Case One: High Scorer (Chen)

Participant 28 (Chen) was an advanced reader who scored highest on both reading comprehension tests. Her scores on the familiar text (language teaching) and unfamiliar text (business) were respectively 10 and 9. Chen employed 142 total strategies with the familiar text and 145 strategies with the unfamiliar text.

Chen's general approaches to the iBT reading comprehension section were to read the entire passage first, and she was the only high scorer in this study who used this reading process. During Chen's interview, she mentioned that although she read through the entire passage before approaching the multiple-choice questions, she did not read the

text in a very detail-oriented way because the multiple-choice questions later would help to identify the key points of the passage. For Chen, reading the passage first was for the purpose of getting an overall understanding of the text, which enabled her to go back and forth between test questions and text to look for an answer. While reading to get an overall gist of the passage, Chen described drawing a clear framework of the text's structure in her head. This reading process was related to the strategies in category 2: drawing a framework of the passage by identifying important information by the discourse structure of the passage. As Chen stated in the interview:

我覺得我看文章有個優勢，就是我在閱讀時很清楚文章的框架。例如這篇文章一開始是在講什麼，然後接下去是什麼 ...

I feel that I have the advantage to be able to draw the framework in my mind of the article, as long as the article is not too hard. I would have a relatively clear structure in my mind of something like the organization of the article, for example, what this article talks about at the beginning, and then the following part etc.

After building a framework of the text's discourse structure, Chen continually took advantage of the multiple-choice testing format. This could be seen from the strategies related to the multiple-choice test-management category with a total of 127 usages on both familiar and unfamiliar texts. For Chen, constructing the textual framework in her mind was more useful in conjunction with the strategies related to multiple-choice test questions. For example, when reading multiple-choice questions, Chen constantly remembered the part of the text that corresponded to the test questions. That is, Chen's mental framework helped her correctly identify the location of the text covered by the multiple-choice questions and, thus, she rarely needed to search the text from the beginning. For example, in question 5, Chen identified the key words as "teachers worried..." and she was able to instantly connect "concern" and the related passage in the text.

我一看完題目後，就想起文章有這個同意字可以對應，所以就回去文章 那部分再字細看一次，其他地方就不用再浪費時間了。

I remembered there was a vocabulary "concern" somewhere in the text that I need to find out, since it was talking about "teachers' worries" in the question. So, I

easily located the specific text part and reread those related sentences to find out the best answer.

Through Chen's effort in framework building as well as her use of strategies related to multiple-choice test questions, she demonstrated a high level of metacognitive awareness by recognizing the simulated iBT reading task and adopting reading strategies different from non-testing situations. For example, relying on her background knowledge was not her primary test-taking strategy because she knew the test questions would lead the focus. Even when the content area was familiar for Chen, she thought it was important to understand the test author's emphasized details as reflected by the questions on that specific passage. In addition, although sometimes Chen encountered reading problems, she would skip those unclear parts because they did not interfere with her catching the main idea, and she would go back to read them later to clarify meaning if the questions asked. For example, she skipped one inferencing question and returned to it after all the other questions were completed. As she noted in her interview:

我通常考試閱讀都以整體概念為主，遇到不太清楚的地方就先跳過去。因為很可能最後題目根本也沒有問到那部份，最重要的是把整體文章搞懂，萬一題目問到，再回來看仔細，不會的題目也就先跳過。

I usually read in order to get the large picture of the text. Sometimes comprehension problems occur, don't waste too much time on that and you may just skip it. It does not really matter whether you understand it or not, because there may be no questions asking about it. For me, it's more important to let the process go. Also, skip those difficult questions and make sure you complete those easy ones first.

Chen's mental framework of the passage facilitated the process of going back and forth between test questions and the text in order to reread for specific information, particularly while answering inferencing questions related to sentence level or passage level understanding. With basic comprehension questions, such as vocabulary synonyms, Chen had a very high level of vocabulary knowledge, which made these questions easy for her, but she still paid attention to the information in the context, as she employed 28 strategies related to vocabulary-in-context approaches. For example, she confirmed the

meaning of “objective” with the neighboring phrase “standing outside of the situation” before selecting her response.

因爲出題者一定會給予一些同意字的線索，大部分是在前後句子裡面，所以就算我很確定某個單字的意思，我還是會去確認一下。

From the perspective of the question designer, there has to be a context based on which we can find out the meaning of the word. Since I was fairly familiar with the vocabulary meaning, I realized that ‘unbiased’ should be the right one.

Chen focused on understanding the big picture of the text in the reading comprehension test, and she was able to comprehend the passage successfully. In Chen’s case, test-wiseness strategies were used only four times. These processes of elimination and guess strategies were not randomly used and were usually a last resort. That is, these strategies were adopted only when her comprehension of the text did not lead her to the best response selection. Because understanding the passage was Chen’s primary concern when taking the reading comprehension tests, the most effective iBT test preparation for her was to memorize the vocabulary words as much as possible in order to comprehend the text successfully.

Case 2: High Scorer (Tai)

Participant 21 (Tai) was also an advanced reader who scored high on both reading comprehension tests but with different test-taking processes than Chen. His scores on the familiar text (law) and unfamiliar text (engineering) were 10 and 8, respectively. He employed 113 total strategies on familiar text and 112 strategies on unfamiliar text.

Tai’s general approach to the simulated iBT reading test was to read the topic sentences in each paragraph to get a general idea of the passage, and then he started to read the questions in order. Reading the questions prior to the text demonstrated taking advantage of iBT reading comprehension questions, which he revealed in his interview, was because each question explicitly pointed out which paragraph it was referring to. Also, the sequence of the iBT questions mirrored the paragraph’s content. Thus, similar to most of the participants, Tai segmented the whole passage into small chunks and tried

to identify the key points in the text from each question. From his interview response:

我先看題目，而且就算做完整篇閱讀，也不會把文章每個單字從頭到尾看一遍。因為考試時間有限，對我來說，考托福閱讀就如同訓練未來當研究生一樣，要學會在最短時間內，想辦法找出重點。

I read the iBT questions first and I don't bother to read every single word in the text because the time is limited in test-taking. I think taking the iBT test is like training us for attending graduate school, which requires us to get the main points in a very short amount of time.

While reading each multiple-choice question, Tai would make a mental note of the key points of the question and jump immediately to where the question was answered in the text. As far as basic comprehension questions were concerned, Tai had a very high level of vocabulary knowledge, which enabled him to correctly select synonyms. When facing unknown vocabulary, he applied the process of elimination first, and then looked for semantic clues from the neighboring sentences to infer meaning. He adopted similar strategies when completing the two reading tasks, but for the familiar text, Tai's related background knowledge about law helped on items for which other test-takers could not infer the correct answer. An explanation was provided in his think-aloud response:

我不知道“preponderance”這個字的意思，但是我從上下文我可以推論出這邊是在講刑事訴訟和民事訴訟的差別。不像刑事訴訟，民事訴訟只需要大部分的證據就可以，所以我知道是選“majority”這個字。不可能是“lack”，因為沒有證據不可能打贏官司。

I don't know this word “preponderance,” but from the text I know they are talking about there are some differences between criminal cases and civil cases. From my background knowledge, civil cases can be decided on only the “majority” of evidence, unlike criminal cases. You can't select “lack”, because you can't win if lack of evidence.

Despite his background knowledge in law, Tai still focused on comprehending the text mainly through multiple-choice test-management strategies. When a mental note of the question was made and the text portion was correctly identified, Tai read the related text part carefully and paraphrased the option's meaning with the text. As far as paragraph-level questions were concerned, Tai was able to recognize how different discourse parts of the paragraph interrelated, such as the introduction, examples, and

conclusion. In this way, Tai was capable of extracting the key idea from a paragraph by referring to the questions. One example follows:

這題問說爲什麼被告者寧願是涉入民法，而非刑法？所以我就可以推斷出，這整段的重點就是“incarceration”個字。因爲刑法需要坐牢。

This question asked why defendants preferred being convicted in civil cases instead of criminal cases? After reading this portion of the paragraph, I could infer the key word as “incarceration”, because criminal cases required jail time.

It seemed from the interview response that Tai devoted his attention to the passage understanding regardless of his background knowledge about the reading passage. He adopted the question-directed reading approach to reach a global comprehension of the passage, and his test-taking processes could be characterized as a repetitive back and forth between the multiple-choice questions and text rereading. The strategies related to background knowledge and test-wiseness categories were only executed if Tai could not make the best selection based on the passage.

The sequence of reading the multiple-choice questions or the passage first in a reading comprehension test demonstrated Tai’s metacognition awareness during the test-taking situation. As he mentioned in the interview:

我會依照文章長短來決定先後順序。假設文章很長我會先看題目，因爲我以前曾經先看文章才看題目，但在看題目時就忘記了文章內容，還得回去文章中找。假設文章很短，我可能就會剛開始從頭把他看完。

Depends on the length of the article. If the article is long like the iBT, I’ll read the questions first. I used to read the whole article; after that, I started the questions but then I had to go back to the paragraphs again and I forgot most of the information. If the article is short, maybe I’ll read the whole article first then answer the questions.

Given that the iBT reading passages are quite long and time is limited, Tai employed the question-directed reading procedure which made him more adept in identifying the main points in the text and allowed him to skip the text not covered by the questions. Tai stressed developing his strategies through cumulative iBT mock exam test practice. Although he adopted different test-taking processes than Chen, both showed that

successful readers, as inferred by their high tests scores, are good strategy users, and they know how to use a variety of goal-oriented, planned strategies.

Case 3: Low Scorer (Gao)

Participant 14 (Gao) was a low scorer on both iBT reading comprehension tests. Her scores on the familiar text (business) and unfamiliar text (law) were respectively 3 and 4. Gao employed only 52 total strategies on the familiar text and 58 strategies on the unfamiliar text; most of them belonged to the categories of general reading approaches and multiple-choice test-management. The number of strategies used was much fewer in part because she was unable to provide a clear think-aloud protocol, requiring a great deal of the cuing from the researcher.

Gao's general approach to the iBT reading comprehension test was to read the entire passage first and then answer the multiple-choice questions one by one, going back to the text looking for specific information if necessary. Unlike the high scorer Chen, Gao read the whole passage word by word and answered the test questions in order. As the iBT reading passages increased in length and Gao's insufficient linguistic knowledge for the task was evident, she spent most of her testing time (17 out of 20 minutes) in decoding the vocabulary, and thus insufficient time was left for answering multiple-choice questions. Gao's invariant reading processes and her inflexible strategy choice when encountering comprehension problems revealed her lack of metacognitive strategy use. Without this metacognitive orchestration during the test-taking situation, she was unable to complete the test satisfactorily.

As seen from Gao's interview responses, the lack of vocabulary knowledge caused a major problem, which led to a slow reading speed and a lack of comprehension. Without sufficient time left for answering the multiple-choice questions, Gao was unable to identify the key points from the test questions or to reread certain text portions covered by the questions. Instead, she selected most of her answers based on memory from her first reading and did not go back to the text again. Therefore, many options were selected

mainly because they contained the same or similar vocabulary from the text. Decoding the text in isolation and trying to match the key word from her memory of the text with the multiple-choice questions seemed to dominate the entire test-taking situation. As the repeated vocabulary sometimes was the test developers' intention to identify those test-takers who did not search for contextual support, Gao rarely answered correctly with this strategy.

我看完題目後就直接作答，通常不會再回文章去看。除非那題目中有明顯提示是哪個句子，如果沒有提示在文章中哪裡，我就翻成中文，反正很少回文章再看一遍。

I read the question and chose the answer directly, without going back to the text again because of running out of time. I only go back to the text if the questions explicitly stated where the related sentences are in text. For these questions which didn't mention clearly, I try to translate that into Chinese, so I can answer immediately without going back to the text again.

Insufficient vocabulary knowledge and text comprehension impacted Gao's decision to adopt more strategies related to her background knowledge and test-wiseness. Gao noted in her interview that she relied on her background knowledge if her comprehension did not support the selection of the most appropriate option. Although Gao tried to connect the text with her background knowledge, she rarely monitored or evaluated her predictions, nor did she modify her predictions based on the text.

我選這個答案是因爲，我以前在經濟課有讀過類似的內容。因爲時間不夠，我都只看文章一次然後去聯想我以前的背景知識。

I chose the answer because I had read the similar content before in my economics class. I just read the passage once due to insufficient time and tried to connect to what I have learned in class before, which saved more time since I did not need to read the text again.

In addition, Gao adopted the process of elimination, and in doing so, sometimes she deleted the correct answer. In the end when the time was almost up, Gao came up with wild guesses and selected choices for no apparent reason. Gao was either unable to comprehend the reading passages or misunderstood them because of decoding, language, and vocabulary difficulties.

Case 4: Low Scorer (Tzeng)

Participant 25 (Tzeng) was another low scorer on both iBT reading comprehension tests. Her scores on the familiar text (engineering) and unfamiliar text (language teaching) were respectively 2 and 3. Tzeng employed only 50 total strategies on the familiar text and 53 strategies on the unfamiliar text. Strategies related to identification of English discourse structure and vocabulary/sentence-in-context approaches were rare in Tzeng's processes.

Tzeng's general iBT reading approach was to skip the whole passage and read the multiple-choice questions first. Because there is critical time limit in the iBT reading test, Tzeng was inclined to read only the portion of the passage needed to answer the test questions, which saved her time to finish all of them. Although adopting the same reading approach as Tai, Tzeng's reading scores were not satisfactory because her lack of vocabulary knowledge slowed her reading speed. Without possessing sufficient linguistic knowledge, even the vocabulary synonym selection questions were difficult, not to mention comprehending the text comprehensively. Thus, Tzeng focused on individual words in the questions rather than on sentence level meaning, and she continually searched for these words in the text and then selected the option that contained the information. However, sometimes it was test author's intention to set up a distraction by using similar vocabulary. One example was presented:

這段最後提到 “in the discipline of teaching”，剛好在選項中，有一個提到說 “adopt a more disciplined approach to teaching”。我想只是作換句話說，所以我覺得答案是這個。

The last sentence mentioned about “in the discipline of teaching”, and there happened to be an option that said “adopt a more disciplined approach to teaching...” I thought these two sentences were paraphrases so I chose this one.

Tzeng tried to take advantage of multiple-choice questions to identify the important part of the passage that needed to be read, but she mistakenly selected the wrong part in the text as relevant or was unable to achieve understanding. Under such circumstances, besides her word-matching strategies, Tzeng selected the option based on

the passage's overall meaning. Her failure to capture the specific meaning of certain parts of the passage was demonstrated by her think-aloud protocol.

這提問說作者爲什麼在三四段末提出這些地名？反正這篇都是在講風力發電，我就選一個選項有風力發電的。

This question asked what can be inferred about the states of North Dakota and South Dakota at the end of paragraph 3 and 4? I think the whole passage is talking about electricity from wind, so I chose the option related to this idea.

Tzeng also adopted more strategies related to test-wiseness than did high scorers in this study when trying to select the most appropriate option. Her think-aloud protocol showed her focus on individual words and her attention to matching similar words between text and options. Without a global understanding of the passage and contextual support, Tzeng ineffectively applied some strategies that were detrimental to her reading test performance. For example, guessing was employed when Tzeng was running out of time or unable to comprehend the passage. As reported in her interview response, Tzeng expressed her concern about insufficient vocabulary knowledge and the importance of practicing mock iBT exams as much as possible to familiarize herself with the vocabulary and test management strategies.

Participants' Perception of Test Questions

In the final interview question, participants were asked to indicate the easiest and most difficult test questions, from both familiar and unfamiliar texts, with explanations of their responses. Nine out of 12 high scorers identified the inferencing questions to be the most difficult. These questions focused on sentence-level information as well as on abilities related to connecting information within the text and recognizing the purpose of the text. One example is presented below:

對我來說最難的題型是插入句，因爲我必須先完全了解題目句，然後再去閱讀相關要被插入的文章地方。而且當在閱讀文章的時候，還要須了解他的細節和整體的邏輯概念，整合出整段落再討論什麼。

The question that asks me to insert a certain sentence into a paragraph is the most difficult one. Because I first need to understand the whole sentence and then I read the article. As I am reading the article, I get some detailed facts and logics of it, and according to that, I try to put the sentence into the article logically. The

difficult part lies in that you need to be able to integrate what the whole paragraph or text is talking about.

Only three high scorers out of 12 identified basic comprehension questions, which focused on the meanings of individual words or phrases, to be the most difficult. This happened especially when these specific participants did not know the meaning of vocabulary words, and the context was not clear enough to help infer meaning.

對我來說最難的是單字題，尤其當我不認識這個單字，而又無法從上下文推論他的意思。句子推論題至少還是可以從上下文來判斷。

The most difficult one for me is the question that asks about the meaning of words, especially some of them don't give me any clue. But for the inferencing questions, I can always guess the meaning from the neighboring sentences, which is easier to me.

The low scorers tended to view both basic comprehension and inferencing questions to be difficult. Basic comprehension questions were difficult because the participants' insufficient linguistic knowledge hindered their ability to correctly identify the corresponding synonyms. In addition, inferencing questions were complicated because low scorers were unable to get the general meaning of the text and to integrate its different parts as a whole. An example is provided below:

對我來說，推論題很難但是單字題也不簡單。因為推論題你必須知道作者在講什麼，而單字題我常常遇到不認識的單字。

As far as I am concerned, both inferencing and basic comprehension test questions were difficult. Inferencing questions required me to understand the author's opinions and to further infer the meanings. Basic comprehension questions were hard because there were so many unknown vocabulary words.

Summary

This chapter reported the results from both the qualitative and the quantitative analyses of the collected data. As shown at the beginning of the chapter in Table 6, six broad categories of 39 strategies were identified. In terms of the quantitative findings, the types of test-taking strategies adopted by Chinese-speaking graduate students at the University of Iowa were similar when they read familiar versus unfamiliar topics. However, the employment of test-taking processes differentiated between high-scoring and low-scoring readers. The next chapter discusses the major findings in this study.

CHAPTER 5

DISCUSSION OF FINDINGS AND CONCLUSION

In this chapter, I present the summary of this study, which includes the main findings from two perspectives: the comparison of test-taking strategies with texts of different topic familiarity and with participants' scores on the reading comprehension tests. Then, I discuss issues that emerged from this study related to the research questions as well as implications for pedagogy and future research.

Summary of this Study

The purpose of this study was to examine, through think-aloud protocols and interviews, the test-taking strategies adopted by Chinese-speaking participants at the University of Iowa when responding to multiple-choice reading comprehension questions based on the simulated TOEFL test. Twenty Chinese-speaking international students completed three tasks: a topical knowledge vocabulary assessment; two practice iBT reading comprehension tests, one with a familiar topic and the other with an unfamiliar topic, both with retrospective think-aloud protocols; and an interview related to their reading patterns and test-taking strategies.

Based on participants' think-aloud protocols, six categories of strategies emerged in this study: general approaches to reading the passages, identification of important information by the discourse structure of the passages, vocabulary/sentence-in-context approaches, multiple-choice test-management strategies, test-wiseness, and background knowledge. Those six categories of strategies are discussed according to participants' different topic familiarity and test scores.

High scorers generally tended to view the basic comprehension questions on the iBT reading tasks, which required them to select the synonym of a vocabulary word or the meaning of a sentence, as the easiest questions. The most challenging questions for high scorers were the inferencing questions, whereas low scorers tended to report both types of questions as challenging.

Strategies Compared with Topic Familiarity

The results of this study revealed a number of interesting patterns in test-taking behaviors. For the purpose of performing well on the reading comprehension test, participants were focused on finding the best answer for each multiple-choice question. Although each participant employed a combination of specific overall strategies in his or her approach to the reading comprehension task, the common behavior that directed the participants was to interpret the questions as quickly as possible and then to use the questions to guide their search of the passage to locate the potential answers. Regardless of whether the passage or the multiple-choice questions were read first, participants constantly focused on a search for information in the passage to answer the questions.

The results of this study revealed that the types and frequency of use of test-taking strategies by Chinese-speaking students were not different when they read familiar versus unfamiliar texts. Although similar kinds of strategies were used with both topic types, unfamiliar topics prompted a modest increase in test-taking strategy use in this study, but not in a statistically significant way. Participants focused more attention on processing language skills when reading unfamiliar texts than they did when reading familiar texts.

Although background knowledge did not influence participants' test-taking processes, it impacted test-takers' affect more than their reading patterns. In the interviews, participants all revealed feeling more relief and more confidence when reading the text related to their background knowledge. They felt more self-assured when reading materials were familiar, especially in a high-stakes testing situation.

Strategies Compared with Test Scores

The high- and low-level groups of participants were compared on their performance in the combined test scores of the two iBT practice reading tests. In terms of strategy use, the statistical analysis showed a strong positive relationship with participants' performance and their strategy use in the first four broad categories: general approaches to reading the passages, identification of important information by the

discourse structure of the passages, vocabulary/sentence-in-context approaches, and multiple-choice test-management strategies. The statistical analysis also showed that test-wiseness strategies were used significantly more frequently by the low-performing group, and the strategy use related to background knowledge was not statistically significant between high- and low-level groups.

High scorers were more aware and more able to report their use of strategies even in their first language. Conversely, low scorers used fewer types and numbers of strategies and usually resorted to guessing in the end. Even if the same type of strategy was used by both groups, the low scorers differed from the high scorers in that high scorers deployed those strategies to fit the needs of the contextual environment. For instance, low scorers tried to match key words from the options with the text. They did not orient their reading based on contextual support and relied only on individual words, which focused on isolated information.

Discussion

Based on the findings of this study, the types of test-taking strategies adopted by Chinese-speaking graduate students at the University of Iowa remained similar when they read familiar versus unfamiliar topics. However, the employment of test-taking processes differentiated between high-scoring and low-scoring readers. Several compelling findings that emerged from this study are discussed below.

Similar Strategies Across Topic Familiarity

In terms of topic familiarity, participants in this study used similar strategies, regardless of the text's content, when completing the reading comprehension test. This finding contradicts previous research that found significant differences when students read texts with different topic familiarity (Peretz & Shoham, 1990; Pritchard, 1990; Racht & Leslie, 1988). One possible explanation is that participants in this study were required to complete simulated high-stakes TOEFL reading comprehension tests. Given that different reading tasks distinguish readers' strategy selection (Singhal, 2001; Phakiti,

2003), the results of this study might be different from results of studies conducted in non-testing reading situations.

It is possible that students in both the high-scoring and the low-scoring groups in this study employed a fixed set of test-taking strategies that they learned from their reading teachers or coaching schools when preparing for the TOEFL test (Huang et al., 2006). As Alderson (2000) argued, reading is a purposeful activity. Participants developed their strategies as the most efficient way to obtain high scores according to their previous experiences when preparing for or taking the TOEFL iBT reading comprehension test. Under such circumstances, students had become accustomed to their overall test-taking techniques without making adjustments for topic familiarity, which made the reading process that they used in comprehending both texts constant.

The use of similar strategies across topic familiarity reflected the assumption of a background knowledge threshold (Clapham, 1996), which claims that only medium-level scorers are most affected by background knowledge. High scorers are able to perform like native speakers who do not rely as heavily on their background knowledge, whereas low scorers cannot take advantage of their background knowledge because they are too focused on vocabulary decoding. This study's focus only on high and low scorers may explain the lack of influence of background knowledge on comprehension.

In addition, unlike general reading textbooks, which are designed for readers to relate the materials to their background knowledge as well as to rely appropriately on their prior knowledge and the text, the TOEFL reading test avoids building barriers based on test-takers' lack of background knowledge. From this perspective, the TOEFL is not supposed to be advantageous for some people and disadvantageous for others based on their background knowledge. Therefore, the TOEFL is designed so that topic familiarity does not influence reading comprehension processes and performance.

Although similar kinds of strategies were found with both topic types, participants exercised a modest increase in test-taking strategy use when reading unfamiliar texts, but

not in a statistically significant way. This finding of more frequent use of processing language skills when reading unfamiliar texts echoed Alsheikh's study (2011) that readers used more strategies when approaching more difficult texts. The possible explanation is that test-takers may not have related background knowledge about the text with which they are unfamiliar, so they need more strategies to process the text.

Different Strategies Across Test Score

The major difference between high and low scorers in this study was in the four category of strategies used. This finding supports previous research that there are clear distinctions between the strategies that proficient and less proficient readers use, as discussed in Chapter 2 (Alderson, 2000; Block, 1992; Huang et al., 2006; Phakiti, 2003). The employment of different strategies by high and low scorers is also related to differences in metacognitive awareness and differences in formal knowledge application.

Differences in Metacognitive Awareness

High scorers and low scorers in this study differed in the first category of strategies used, general approaches to reading the passages. When approaching the reading comprehension test, all participants demonstrated test-oriented reading behavior as found in Farr et al. (1990) and Tian's (2000) studies. Regarding reading comprehension tests as a special kind of reading task, test-takers were constantly driven by a specific reading purpose: to find the answer for each multiple-choice question. That is, as opposed to normal academic reading or reading for personal interest, the task of taking a reading comprehension test requires that test-takers' most attention be devoted to getting the correct answer as quickly as possible.

In this regard, the overall approaches in the first category fall into the realm of metacognitive regulation, as those strategies are executed based on participants' analysis of the specific iBT task. Therefore, the differences in strategies used in the first category by both high and low scorers reinforced the findings from previous research that effective readers coordinate strategy use with metacognitive knowledge (Anderson, 1991; Block,

1986; Cohen, 1998; Phakiti, 2003; Singhal, 2001) and that successful readers demonstrate their overall metacognitive awareness of their reading process. High scorers are generally more aware of the cognitive demands of the given task. For example, although high scorers finished reading the whole passage while taking the reading test, they seemed to be more concerned with the overall meaning of the passage and skipped the supplementary points, instead of devoting their attention and testing time in trying to understand every single word in comparison to low scorers.

Results showed that participants who reported more awareness of metacognitive strategy use during the reading process, such as monitoring their reading comprehension, adjusting their reading rates by calculating the remaining time, and focusing their attention on where the potential answer could be, tended to be high scorers. For example, one high scorer in this study had no previous experience with taking the iBT reading test. She reported that she read the whole passage first before approaching the multiple-choice comprehension questions when performing on the first iBT test. However, in the second iBT test, she read the multiple-choice questions first before reading the text as she realized that the questions helped to guide her reading and more attention could be devoted to the text covered by the questions. As shown by the interview responses, other high scorers in this study revealed their experiences of trying to determine the most appropriate test-taking strategies when they were preparing for the iBT test. As a result, high scorers demonstrated the importance of experimenting with different strategies to find out which were most beneficial instead of strictly following certain recommended strategies.

Studies in L2 reading have come to emphasize the role of metacognition. As Anderson (1991) argued, “strategic reading is not only a matter of knowing what strategy to use, but also the reader must know how to use a strategy successfully and orchestrate its use with other strategies” (p. 468). In this study, high scorers were able to monitor the effectiveness of their reading patterns and apply fix-it strategies when completing the

reading tasks. However, low scorers, who tended to lack this awareness, decoded the whole passage from the beginning without recognizing their reading purpose and were left with insufficient time for answering all of the multiple-choice questions.

Differences in Formal Knowledge Applications

High scorers also demonstrated their awareness of English textual structure, as presented in the strategies of identification of important information by the discourse structure. By using their knowledge of the hierarchical structure of English, high scorers took advantage of reading the topic sentences, which usually represent the main idea of a paragraph, as well as identifying transition words, which function as logical connectors to facilitate building meaning-making.

As Lin (2002) discussed, these “pointer words or signals” indicate sequential or causal relationships in text structures. Previous studies have shown the importance of knowledge of textual organization in L2 reading (Bernhardt, 1991; Carrell, 1984), which corroborates the significance of formal knowledge, as discussed in Chapter 2. The significant difference between the strategies used in this category by high and low scorers may be due to the differences between Chinese and English language structure. Although English writing usually conveys main points through topic or concluding sentences in each paragraph and interrelates different parts of the text through transitional connectors, Chinese passages usually follow the traditional four sequences: introduction, elucidation of the theme, transition to another viewpoint, and summing up at the end. Thus, low scorers in this study may have lacked the formal knowledge of English rhetorical patterns and thus were unable to use this formal knowledge to facilitate their comprehension. This result supports that having a formal schema aids readers in completing a reading task well (Singhal, 1998; Stone, 1985).

The Influence of Chinese Pedagogical Cultural Background

It has been acknowledged that readers’ pedagogical cultural background influences their comprehension strategies because readers have learned to read within the

context of specific cultures (Abbott, 2006; Block, 1986; Parry, 1996). Therefore, it is likely that Chinese EFL readers' L1 processing skills inevitably affect their strategies when reading an English text. With the strategies Chinese students have developed through exposure to specific Chinese literacy practices, these EFL learners tend to utilize the bottom-up reading strategies and encounter problems with integrating an English text as a whole (Kohn, 1992). This pattern emerged with the low scorers in this study. However, the high scorers were more able to adjust their strategy use by increasing global awareness of the text to enhance comprehension (Block, 1992; Tian, 2000). Low scorers focused more on word-level decoding skills without paying attention to synthesizing information as a whole. Therefore, in order to best prepare readers to comprehend successfully, balanced reading curricula that emphasize both bottom-up and top-down reading skills and strategies are necessary to help readers from different linguistic and cultural backgrounds to be more successful, as these curricula would allow readers to capitalize on their strengths (Abbott, 2006).

The Impact of L2 Vocabulary on Test-taking Strategies

The strategies employed by and interview responses from the low-scoring participants provided evidence that they were lacking sufficient vocabulary knowledge to comprehend the passages. Researchers have attributed unsuccessful readers' comprehension difficulty to their lack of lower-level decoding subskills (Lin, 2002), despite their dependence on bottom-up strategies. Without passing a basic vocabulary threshold level, there is a "short-circuit" effect (Clarke, 1988) for those participants whose linguistic proficiency is too low to make efficient reading possible. That is to say, low scorers in this study faced the difficulty of vocabulary understanding and, therefore, had to spend most of their time in decoding the text from the beginning to arrive at a basic understanding of the text.

As discussed above, limited vocabulary knowledge negatively impacts low scorers' selection of strategies and negatively affects their test performance. Low scorers

ineffectively applied some strategies, although with the same intention as high scorers. For example, to make inferences about vocabulary and sentences in context, low scorers tended to match the text with the options without the support of contextual clues. High scorers tended to translate the overall meaning from sentence to sentence, whereas low scorers mostly focused on individual words. Although low scorers were aware of the importance of finding key information in the text, their insufficient vocabulary knowledge hindered them from distinguishing main ideas from minor examples, which made them unable to identify the key points and to skip some portions of the text.

Studies have demonstrated that successful reading requires constant interactions between bottom-up vocabulary decoding and top-down meaning-making processes (Nikolov, 2006). As shown by the qualitative results of this study, the high scorers demonstrated their attention to both processes by combining strategies related to word-level reasoning processes, such as strategies in the category of vocabulary/sentence-in-context approaches; with top-down predicting processes, such as strategies related to the use of background knowledge, to reach a comprehensive understanding of the text.

Low scorers in this study obviously increased their strategy uses in the category related to background knowledge. This can be explained by the interactive-compensatory model (Stanovich, 1980). According to this model, reading is an interactive process that requires readers to use both top-down and bottom-up processing, and having additional information about one aspect can compensate for lacking information from another aspect. In this regard, background knowledge helped low scorers in this study with top-down information, which can compensate for the information they could not acquire from their bottom-up processing due to insufficient linguistic knowledge. However, the compensation was not enough to increase their test scores. As L2 linguistic knowledge is of crucial importance to beginning learners as well as for those whose L2 competence has not reached the vocabulary threshold level (Clarke, 1988), low scorers in this study

ultimately adopted more background knowledge, which they assumed would help in comprehension.

As Carrell (1988) argued, readers experiencing problems with the words or structures in the text might put these problems aside and instead devote more of their attention to making hypotheses about the text's content based on their background knowledge or personal experiences. Thus, it seems that low scorers in this study reported relying on their background knowledge as much as possible when they encountered reading problems caused by insufficient vocabulary knowledge. Unlike low-scoring readers who primarily used their background knowledge, high-scoring readers related and integrated text information with their background knowledge and effectively continued to confirm meanings, to make corrections, and to adopt strategies to construct meaning in the text.

High scorers perceived sufficient vocabulary knowledge, which enabled them to read fluently and attend to the meanings of the iBT text without being slowed down by word-recognition demands. Word-recognition fluency, a crucial component of reading fluency, can be an influential factor in distinguishing L2 reading comprehension abilities because it is related to a reader's ability to read rapidly with ease and accuracy (Koda, 2005). Segalowitz (2000) argued that word-recognition fluency can be developed through extensive repetition of vocabulary decoding and practice reading. This demonstrates that most of the high scorers in this study had been enrolled in their graduate studies longer than low scorers. This phenomenon explained that the exposure to academic reading in graduate schools improved students' word-recognition performance in both speed and accuracy. On the other hand, low scorers were deficient in word-recognition fluency, as reflected in their difficulty of understanding what they read and an inability to read smoothly.

Insufficient linguistic knowledge that caused a major problem when comprehending the text was also revealed by low scorers during the interview questions

related to which test questions were perceived as the easiest and most difficult by each participant. As far as high scorers were concerned, the basic comprehension questions on the iBT reading task, which required them to select the synonym of a vocabulary word or a sentence's expression of meaning, were perceived as the easiest questions perhaps because they had passed the vocabulary threshold and had sufficient linguistic knowledge. In this case, high scorers perceived the most challenging questions to be inferencing questions. Such questions include test-takers' demonstration of understanding paragraph-level information, to comprehend an argument that was not explicitly stated in the text, and the ability to recognize and engage with the whole text's organization to successfully insert a new sentence in the passage. As far as low scorers were concerned, they tended to report both types of questions as challenging.

Insufficient linguistic knowledge also reduced low scorers' reading speed and lead to a lack of completion of their reading tests. This is a concern because TOEFL iBT reading passages are longer than in previous TOEFL tests. As revealed in their think-aloud protocols and interview responses, low scorers felt that they did not have enough time to read the text and to finish all of the multiple-choice reading comprehension questions. With their limited vocabulary and deficient reading fluency, low scorers needed extra effort and time in their processing of lower-level vocabulary decoding to distinguish main ideas from minor details or examples; therefore, they assumed they could skip the parts that the questions did not address. This statement also corroborates metacognitive strategies discussed before as low scorers are incapable of monitoring their effective of strategy use and applying remediating actions when problems are detected. Participants reported that taking practice mock exams, which help test-takers with time management and reading speed, as often as possible before taking the real test was the most effective test preparation method.

Multiple-choice Test Management Strategies

Rupp et al. (2006) argued that different testing formats result in different strategy use; thus, multiple-choice tests create a unique testing format and elicit specific strategy selection. For example, multiple-choice questions enable test-takers to get clues about the important parts of the text and to ignore the rest of the parts not covered by the test questions. Unlike other reading comprehension tests that require free responses or summary recall, participants are aware of the advantages of the multiple-choice format and use the questions and options as indicators of which portion of the text tends to be more important or worth reading, especially under critical time constraints.

Multiple-choice test management strategies were the most frequently applied in this study. Such strategies included reading the multiple-choice questions first and then jumping immediately to the text indicated by the questions; making a mental note when reading the questions to guide participants' search while reading; inferring and extracting the key points of the text from the questions or options; rereading or paraphrasing the text with the options; and selecting the option because it contains a possible key word that appeared in the text.

Although all participants adopted multiple-choice test management strategies, it seems that high scorers used these multiple-choice strategies in combination with comprehension. That is, high scorers selected their answers mostly based on their understanding of the text. For instance, high scorers reported that they made a mental note of reading the questions and using them to guide their search while reading, and when they correctly identified the text where the question appeared, they devoted their attention to trying to understand that portion of the text. With their understanding of the text, they were able to attend to higher-level comprehension processes; to distinguish the main points with supplementary details; and to skip those parts of the passage that were not covered in the questions. They were also capable of paraphrasing the text correctly given the options available.

However, given that the low scorers adopted a repertoire of strategies that were similar to those of high scorers, the difference in their performance was shown to be in their inability to regulate these multiple-choice test management strategies. As Anderson (1991) argued, it is important to know when and why to employ certain specific strategies and how to employ them successfully. For low scorers, multiple-choice test management strategies were not applied effectively. Because they were unable to identify the text's main points even after reading the questions first, they selected an option because it contained a word or phrase that appeared in the text instead of confirming meaning by context. Low scorers often fell into this trap if they did not search for contextual support and relied only on shared words between the text and the response options.

The results indicated that multiple-choice strategies benefited participants if they were adopted appropriately in combination with comprehension. The reading processes and strategies that participants' demonstrated in this study support the validity of multiple-choice reading comprehension tests because the test-takers selected their answers mostly based on their understanding of the text. Participants in this study viewed questions and passages as one interrelated task, rather than only concentrating on the questions (Farr et al., 1990). This substantiates previous research suggesting that test takers read the passage as much as possible to enhance global comprehension in order to achieve higher scores (Daneman & Hannon, 2001) and appears to contradict previous studies (Cohen, 1984) in which examinees did not read or comprehend the test passages.

Test-wiseness Strategies

In this study, strategies identified in the test-wiseness category overlapped with multiple-choice test management strategies. While multiple-choice test management strategies focused on the use of multiple-choice questions and options, test-wiseness strategies focused on the processes of elimination and guesswork, which included: eliminating other impossible options even when the selected response was not understood; eliminating options not mentioned in the designated paragraph; and choosing an option

based on guesswork because it appears to be the longest answer or the frequency of each answer should be evenly distributed.

Test-wiseness is a test-taker's ability to utilize the characteristics and formats of the test and the test-taking situation to receive a higher score, without necessarily knowing the content or using the skill that test authors intended to test (Allan, 1992; Dodeen, 2008). According to Cohen (1998), test-taking strategies are not inherently effective or ineffective because their successful use relies on whether they are appropriate for a specific task and are properly adopted to complete the task. This study demonstrated that certain test-wiseness strategies were adopted by both high and low scorers on the reading assessment. However, those strategies benefited only the high scorers and increased their test scores because they adopted those strategies more effectively and under appropriate circumstances. For example, although both groups of participants used the process of elimination, only high scorers applied this strategy effectively because they were confident that the rest of the options would not be correct due to different synonyms or contrasts with the passage's overall meaning. Therefore, the process of elimination did not work well for the low scorers because they were unable to identify the incorrect options. Because of their difficulties with comprehending the text, low scorers resorted to guessing or choosing the longest answers if they were running out of time..

It is interesting that test-wiseness was the only category of strategies that low scorers employed significantly more often than high scorers in this study. Low scorers not only applied the test-wiseness strategies discussed above ineffectively, but also used more incorrect strategies to compensate for their lack of understanding of the text. The findings from the think-aloud protocols and interview responses showed that low scorers relied on test-wiseness strategies to compensate for their insufficient linguistic knowledge and failure to understand the text. As mentioned above, readers who experience problems with vocabulary or sentence structures might seek to compensate for these problems by directing more of their attention to other strategies that they think might help. Thus, low

scorers applied significantly more test-wiseness strategies when they encountered problems in determining the best option, which happened frequently.

Even though some participants did not read every single word in the iBT reading passage, the test retained its validity in the broader sense because high scorers focused on the main points based on their comprehension of the text and used test-wiseness strategies sparingly. The discrepancy between high and low scorers' use of test-wiseness strategies might be due to their different comprehension levels based on their reading ability. Therefore, the less frequent use of test-wiseness strategies by high scorers indicated that they viewed test-wiseness strategies as a supplementary device when their understanding of the text was not sufficient to select an answer. The interview responses also supported that high scorers seemed to regard test-wiseness strategies as a last resort, that they played a secondary role in the task of completing the multiple-choice reading comprehension test, and that they were always used in combination with overall text understanding.

Implications for Test-taking and Pedagogy

The findings of this study were based on participants' think-aloud protocols and interview responses that described their test-taking behaviors when completing two practice TOEFL iBT reading multiple-choice reading comprehension tests. One implication of this study is related to the test-taking strategies employed by the participants when completing the reading comprehension tests. That is, the strategies that were beneficial to multiple-choice reading comprehension test performance in general, and to the TOEFL iBT reading comprehension section, in particular were identified and discussed.

This examination of test-takers' reading processes provides evidence for the validity of the TOEFL iBT reading test in measuring reading comprehension, as findings related to how test-takers arrive at their responses in different assessment tasks and contexts provide insights to test validity (Cohen, 2011). The results in this study show

that comprehension is still the fundamental element required for a test-taker to perform well on the TOEFL iBT reading test, although the multiple-choice questions appeared to cue important parts of the texts for test-takers, thus eliciting cognitive and metacognitive strategies that may not occur with other testing formats. The reading tests used in this study retain their validity because high scorers focused on the main points based on their comprehension of the text even though they sometimes applied test-wisness strategies. This finding substantiated the results of Cohen and Upton's (2007) research in that the iBT reading section evaluates test-takers' abilities to demonstrate a combination of academic reading comprehension and test-taking skills to achieve satisfactory results.

Test-wisness strategies and background knowledge should not be ignored, although they were not considered as primary strategies used by high scorers. In test-taking situations, test-wisness and background knowledge are beneficial to test-takers' performance if they are applied appropriately. Test-takers should relate important points in the text to their background knowledge in order to comprehend the text as a whole and to modify their connections according to the text content. Test-takers should adopt test-wisness strategies carefully as they are a last resort.

In addition to the test-taking implications described above, some implications related to reading pedagogy also emerged. The major problem that low scorers encountered was insufficient linguistic knowledge that hindered their reading comprehension. Based on their reports, low scorers processed information at the word level by mainly focusing on a bottom-up decoding strategy. However, high scorers appeared to comprehend the text at a global level, with help from their background knowledge and text structural knowledge to handle any linguistic difficulties (Block, 1992). Given that reading comprehension is more than understanding at the word and sentence levels, reading instructors should stress the importance of whole text understanding. Because the current TOEFL iBT requires higher-level comprehension such as inferencing questions, the ability to integrate a text's macrostructure instead of

focusing on lower-level word processing skills should be emphasized in language learning lessons.

Another implication for reading pedagogy is the importance of teaching students metacognitive awareness during their test-taking situations. Because of the longer reading sections in the current TOEFL iBT test, high scorers tended to break the whole text into smaller segments and to read these chunks aligned with related questions. Even though a high scorer read the whole text once, she tended to construct the framework of the text in her head and to change strategies when problems were detected. Differences between high and low scorers are found in their problem-solving abilities and their implementation of knowing why, when, and how to apply specific strategies according to different reading tasks faced. Reading instructors should help their students be more conscious of their own test-taking strategies, to effectively monitor their comprehension status, and to adopt remediating strategies when necessary (Alderson, 1991; Block, 1992; Tian, 2000).

Limitations and Suggestions for Future Study

Several limitations of this study should be discussed, which may provide improvement for future related studies. This research was conducted with only Chinese language speakers. Thus, any conclusions drawn may be different for speakers from other language backgrounds because readers from different cultures may use different reading strategies. This study relied heavily on participants' think-aloud protocols, a method that is limited since it cannot completely reflect readers' inner processes. This study adopted the paper-based iBT practice reading comprehension test. Although the computer-based language tests and the paper-based language tests are comparable, future research should be completed on-line to more closely approximate the testing conditions of the TOEFL iBT test. This study focused only on two types of TOEFL iBT reading comprehension questions, the basic comprehension and inferencing questions, and did not address the reading to learn questions. Different reading processes may be developed for different

question types, even if these question types are elicited within the same testing format (multiple-choice questions). Future research might be conducted with a greater variety of question types so that other test-taking strategies can be explored.

Conclusion

The results of this study extend the findings of Tian (2000) and Upton and Cohen (2007) by comparing the participants' test-taking strategies with different levels of topic familiarity and with participants' scores on the reading comprehension tests. Although it has been demonstrated that topical knowledge can be a significant factor affecting L2 reading comprehension, this study showed that the strategies used by participants in the test-taking situation to process text were similar regardless of topic familiarity. However, the employment of strategies can differentiate high and low scorers. This study contributed to L2 reading assessment by presenting the test-taking strategies in regard to different topic familiarity and by demonstrating the validity of multiple-choice reading tests.

REFERENCES

- Abbott, M. L. (2006). ESL reading strategies: Differences in Arabic and Mandarin speakers' test performance. *Language Learning*, 5(4), 633-670.
- Alderson, J. C. (2000). *Assessing reading*. New York, NY: Cambridge University Press.
- Alderson, J. C. (2009). Test of English as a Foreign Language™: Internet-based Test (TOEFL iBT®). *Language Testing*, 26(4), 621-631.
- Allan, A. (1992). Development and validation of a scale to measure test-wiseness in EFL/ESL reading test takers. *Language Testing*, 9(2), 101-119.
- Almasi, J. F. (2003). *Teaching strategic processes in reading*. New York, NY: Guilford Press.
- Alsheikh, N. O. (2011). Three readers, three languages, three texts: The strategic reading of multilingual and multiliterate readers. *The Reading Matrix*, 11(1), 34-53.
- Anderson, N. J. (1991). Individual differences in strategy use in second language reading and testing. *The Modern Language Journal*, 75(4), 460-472.
- Andreassen, R., & Braten, I. (2010). Examining the prediction of reading comprehension on different multiple-choice tests. *Journal of Research in Reading*, 33(3), 263-283.
- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice*. New York, NY: Oxford University Press.
- Bang, H. J., & Zhao, G. F. (2007). Reading strategies used by advanced Korean and Chinese ESL graduate students: A case study. *The Reading Matrix*, 7(1), 2007.
- Bernhardt, E. B. (1986). Reading in the foreign language. In B. H. Wing (Ed.), *Listening, reading, and writing: Analysis and application* (pp. 93-115). Middlebury, VT: Northeast Conference on the Teaching of Foreign Languages.
- Bernhardt, E. B. (2005). Progress and procrastination in second language reading. *Annual Review of Applied Linguistics*, 25(2), 133-150.
- Blanton, E. L. (2004). *The influence of students' cognitive style on a standardized reading test administered in three different formats* (Unpublished doctoral dissertation). University of Central Florida, Orlando.
- Block, E. (1986). The comprehension strategies of second language readers. *TESOL Quarterly*, 20, 463-494.
- Block, E. (1992). See how they read: Comprehension monitoring of L1 and L2 readers. *TESOL Quarterly*, 26(2), 319-341.
- Bowles, M.A. (2010). *The think-aloud controversy in second language research*. New York, NY: Routledge.

- Brantmeier, C. (2000). *The relationship between readers' gender, passage content, comprehension and strategy use in reading Spanish as a second language* (Unpublished doctoral dissertation). Indiana University, Bloomington.
- Brantmeier, C. (2005). Effects of reader's knowledge, text type, and test type on L1 and L2 reading comprehension. *Modern Language Journal*, 89(1), 37-53.
- Bridgeman, B., & Rock, D. A. (1993). Relationships among multiple-choice and open-ended analytical questions. *Journal of Educational Measurement*, 30(4), 313-329.
- Brown, H. D. (2001). *Teaching by principles: An interactive approach to language pedagogy*. New York, NY: Longman.
- Campbell, J. R. (1999). *Cognitive processes elicited by multiple-choice and constructed-response questions on an assessment of reading comprehension* (Unpublished doctoral dissertation). Temple University, Philadelphia, PA.
- Carrell, P. L. (1983). Some issues in studying the role of schemata, or background knowledge, in second language comprehension. *Reading in a Foreign Language*, 1, 81-92.
- Carrell, P. L. (1985). Facilitating ESL reading by teaching text structure. *TESOL Quarterly*, 19, 727-752.
- Carrell, P. L. (1987). Content and formal schemata in ESL reading. *TESOL Quarterly* 21(3), 461-481
- Carrell, P. L., & Wise, T. E. (1998). The relationship between prior knowledge and topic interest in L2 reading. *Studies in Second Language Acquisition*, 20, 285-309.
- Chapelle, C. A., Enright, M. K., & Jamieson, J. M. (2008). *Building a validity argument for the Test of English as a Foreign Language*. New York, NY: Routledge.
- Chen, Q., & Donin, J. (1997). Discourse processing of first and second language biology texts: Effects of language proficiency and domain-specific knowledge. *The Modern Language Journal*, 81(2), 209-227.
- Chen, H-C., & Graves, M. F. (1995). Effects of previewing and providing background knowledge on Taiwanese college students' comprehension of American short stories. *TESOL Quarterly* 29(4), 663-686.
- Clapham, C. (1996). The development of IELTS: A study of the effect of background knowledge on reading comprehension. *Studies in Language Testing Series, Volume 4*. Cambridge, UK: Cambridge University Press.
- Clarke, M. A. (1988). The short circuit hypothesis of ESL reading: Or when language competence interferes with reading performance. In P. L. Carrell, J. Decine, & D. E. Eskey (Eds.), *Interactive approaches to second language reading* (2nd ed., pp. 183-205). New York, NY: Cambridge University Press.
- Choi, C., Kim, K. S., & Boo, J. (2003). Comparability of a paper-based language test and a computer-based language test. *Language Testing*, 20(3), 295-320.

- Cohen, A. D. (1984). On taking language tests: What the students report. *Language Testing, 1*, 70-81.
- Cohen, A. D. (2011). The coming of age of research on test-taking strategies. *Language Assessment Quarterly, 3*(4), 307-331.
- Cohen, A. D. (1998). *Strategies in learning and using a second language*. Harlow, UK: Longman.
- Cohen, A. D., & Upton, T. A. (2007). "I want to go back to the text": Response strategies on the reading subtest of the new TOEFL. *Language Testing, 24*(2), 209-250.
- Cooper, J. D. (1986). *Improving reading comprehension*. Boston, MA: Houghton Mifflin.
- Creswell, J. W., & Plano Clark, V. L. (2011). *Designing and conducting mixed methods research*. Thousand Oaks, CA: Sage.
- Daneman, M., & Hannon, B. (2001). Using working memory theory to investigate the construct validity of multiple-choice reading comprehension tests such as SAT. *Journal of Experimental Psychology, 130*(2), 208-223.
- Dechant, E. (1991). *Understanding and teaching reading: An interactive model*. Hillsdale, NJ: Lawrence Erlbaum.
- Dolly, J. P., & Williams, K. S. (1986). Using test-taking strategies to maximize multiple-choice test scores. *Educational and Psychological Measurement, 46*, 619-625.
- Droop, M., & Verhoeven, L. (1998). Background knowledge, linguistic complexity, and second-language comprehension. *Journal of Literacy Research, 30*(2), 253-271.
- Eggington, W. G. (1987). Written academic discourse in Korean: Implications for effective communication. In U. Connor & R. Kaplan (Eds.), *Writing across cultures: Analysis of L2 texts* (pp. 153-168). Reading, MA: Addison-Wesley.
- Ericsson, K. A., & Simon, H. A. (1984). *Protocol analysis: Verbal reports as data*. Cambridge, MA: The MIT Press.
- Erler, L., & Finkbeiner, C. (2007). A review of reading strategies: Focus on the impact of first language. In A. D. Cohen & E. Macaro (Eds.), *Language learner strategies* (pp.187-206). Oxford, UK: Oxford University Press.
- Educational Testing Service. (2000). Computer-based TOEFL. Retrieved from <http://www.ets.org/toefl>
- Farr, R., & Carey, R. (1986). *Reading: What can be measured?* Newark, DE: International Reading Association.
- Farr, R., Pritchard, R., & Smitten, B. (1990). A description of what happens when an examinee takes a multiple-choice reading comprehension test. *Journal of Educational Measurement 27*, 209-226.
- Garner, R. (1987). *Metacognition and reading comprehension*. Norwood, NJ: Ablex.

- Grabe, W. (2009). *Reading in a second language: Moving from theory to practice*. New York, NY: Cambridge University Press.
- Grabe, W., & Stoller, F. L. (2002). *Teaching and researching reading*. London, UK: Pearson Education Longman.
- Green, A. (1998). *Verbal protocol analysis in language testing research: A handbook*. Cambridge, UK: Cambridge University Press.
- Hammadou, J. (1991). Interrelationships among prior knowledge, inference, and language proficiency in foreign language reading. *The Modern Language Journal*, 75(1), 27-38.
- Hill, C., & Larsen, E. (2000). *Children and reading tests*. Stanford, CT: Ablex.
- Huang, H-C., Chern, C-I., & Lin, C-C. (2006). EFL learners' online reading strategies: A comparison between high and low EFL proficient readers. *English Teaching & Learning Special Issue*, (1), 1-22.
- Huang, H-T. (2010). *Modeling the relationship among topical knowledge, anxiety, and integrated speaking test performance: A structural equation modeling approach* (Unpublished doctoral dissertation). University of Texas at Austin.
- Hudson, T. (1988). The effects of induced schemata on the "short-circuit" in L2 reading: Non-decoding factors in L2 reading performance. In P. L. Carrell, J. Decine, & D. E. Eskey (Eds.), *Interactive approaches to second language reading* (2nd ed., pp. 183-205). New York, NY: Cambridge University Press.
- Johnson, P. (1982). Effects on reading comprehension of background knowledge. *TESOL Quarterly*, 16(4), 503-516.
- Johnston, P. H. (1983). *Reading comprehension assessment: A cognitive basis*. Newark, DE. International Reading Association.
- Johnston, P. H. (1984). Prior knowledge and reading comprehension test bias. *Reading Research Quarterly*, 19(2), 219-239.
- Kaplan, R. (1966). Cultural thought patterns in intercultural education. *Language Learning*, 16(1), 1-20.
- Karbalaei, A. (2010). A comparison of metacognitive reading strategies used by EFL and ESL readers. *The Reading Matrix*, 10(2), 165-180.
- Keenan, J. M., Betjemann., R. S. (2006). Comprehending the Gray Oral Reading Test without reading it: Why comprehension tests should not include passage-dependent items. *Scientific Studies of Reading*, 10(4), 363-380.
- Kim, M. (2010). *Adult ESL Korean reader's responses about their reading in L1 Korean and L2 English* (Unpublished doctoral dissertation). University of Arizona, Tucson.
- Koda, K. (2005). *Insights into second language reading: A cross-linguistic approach*. New York, NY: Cambridge University Press.

- Kohn, J. (1992). Literacy strategies for Chinese university learners. In F. Dubin & N. Kuhlman (Eds.), *Cross-cultural literacy* (pp. 113-125). Englewood Cliffs, NJ: Regents.
- Krekeler, C. (2006). Language for special academic purposes (LSAP) testing: The effect of background knowledge revisited. *Language Testing* 23(1), 99-130.
- Kvale, S. (1996). *Interviews: An introduction to qualitative research interviewing*. Thousand Oaks, CA: Sage.
- Lee, J. F. (1986). On the use of the recall task to measure L2 reading comprehension. *Studies in Second Language Acquisition*, 8, 201-212.
- Leeser, M. J. (2003). *Second language comprehension and processing grammatical form: The effects of topic familiarity, mode, and pausing* (Unpublished doctoral dissertation). University of Illinois at Urbana-Champaign.
- Lin, Z. (2002). Discovering EFL readers' perception of prior knowledge and its roles in reading comprehension. *Journal of Research in Reading*, 25(2), 172-190.
- Logie, M. (1995). *An examination of comprehension processes used by readers as they complete multiple-choice tests* (Unpublished doctoral dissertation). Indiana University, Bloomington.
- Malik, A. A. (1990). A psycholinguistic analysis of the reading behavior of EFL proficient readers using culturally familiar and culturally unfamiliar expository texts. *American Educational Research Journal*, 27(3), 205-223.
- Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher* 23(2), 13-23.
- Moore, D. W. (1983). A case for naturalistic assessment of reading comprehension. *Language Arts*, 60(8), 957-969.
- Nevo, N. (1989). Test-taking strategies on a multiple-choice test of reading comprehension. *Language Testing* 6(3), 199-215.
- Nikolov, M. (2006). Test-taking strategies of 12- and 13-year-old Hungarian learners of EFL: Why whales have migraines. *Language Learning*, 56(1), 1-51.
- Orasanu, J., & Penney, M. (1986). Introduction: Comprehension theory and how it grew. In J. Orasanu (Ed.), *Reading comprehension: From research to practice* (pp. 1-10). Hillsdale, NJ: Erlbaum.
- Oxford, R. (1993). Research on second language learning strategies. *Annual Review of Applied Linguistics*, 12, 175-187.
- Ozuru, Y., Best, R., Bell, C., Witherspoon, A., & McNamara, D.S. (2007). Influence of question format and text availability on the assessment of expository text comprehension. *Cognition and Instruction*, 25(4), 399-438.
- Parry, K. (1996). Culture, literacy and L2 reading. *TESOL Quarterly*, 30, 665-692.

- Peretz, A. S., & Shoham, M. (1990). Testing reading comprehension in LSP: Does topic familiarity affect assessed difficulty and actual performance? *Reading in a Foreign Language*, 7, 447–455.
- Phakiti, A. (2003). A closer look at the relationship of cognitive and metacognitive strategy use to EFL reading achievement test performance. *Language Testing*, 20(1), 26-56.
- Powell, J. (1988). *An examination of comprehension processes used by readers as they engage in different forms of assessment* (Unpublished doctoral dissertation). Indiana University, Bloomington.
- Pritchard, P. (1990). The effects of cultural schemata on reading processing strategies. *Reading Research Quarterly*, 25(4), 273-295.
- Purpura, J. (1997). An analysis of the relationships between test takers' cognitive and metacognitive strategy use and second language test performance. *Language Learning*, 47, 289-325.
- Radojevic, N. (2006). *Exploring the use of effective learning strategies to increase students' reading comprehension and test taking skills* (Unpublished doctoral dissertation). Brock University, St. Catherines, Ontario, Canada.
- Recht, D. R., & Leslie, L. (1988). Effects of prior knowledge on good and poor readers' memory of text. *Journal of Educational Psychology*, 80(1), 16-20.
- Resnick, L. B., & Resnick, D. P. (1992). Assessing the thinking curriculum: New tools for educational reform. In B. R. Gifford & M. C. O'Connor (Eds.), *Changing assessment: Alternative view of aptitude, achievement, and instruction*. Boston, MA: Kluwer Academic.
- Ridgway, T. (1997). Threshold of the background knowledge effect in foreign language reading. *Reading in a Foreign Language*, 11, 151-168.
- Rodriguez, M. C. (2003). Construct equivalence of multiple-choice and constructed-response items: A random effects synthesis of correlations. *Journal of Educational Measurement*, 40, 163–184.
- Rumelhart, D. E. (1985). Toward an interactive model of reading. In H. Singer & R B. Ruddell (Eds.), *Theoretical models and processes of reading* (3rd ed., pp. 722-750). Newark, DE: International Reading Association.
- Rupp, A. A., Ferne, T., Choi, H. (2006). How assessing reading comprehension with multiple-choice questions shapes the construct: A cognitive processing perspective. *Language Testing*, 23(4), 441-474.
- Saricoban, A. (2002). Reading strategies of successful readers through the three phase approach. *The Reading Matrix*, 2(3), 1-16.
- Segalowitz, N. (2000). Automaticity and attentional skill in fluent performance. In H. Riggenbach (Eds.), *The handbook of second language acquisition* (pp. 382-408). MI: University of Michigan Press.

- Shapiro, A. M. (2004). How including prior knowledge as a subject variable may change outcomes of learning research. *American Educational Research Journal*, 41(1), 159-189.
- Shohamy, E. (1984). Does the testing method make a difference? The case of reading comprehension. *Language Testing*, 1, 147-170.
- Singhal, M. (1998). A comparison of L1 and L2 reading: Cultural differences and schema. *Internet TESL Journal*. Available from <http://iteslj.org/Articles/Singhal-ReadingL1L2.html>
- Smith, F. (1986). *Reading*. Cambridge, UK: Cambridge University Press.
- Stanovich, K. (1980). Toward an interactive-compensatory model of individual differences in the development of reading fluency. *Reading Research Quarterly*, 16, 32-71.
- Sawaki, Y., Stricker, L. J., & Oranje, A. H. (2009). Factor structure of the TOEFL Internet-based test. *Language Testing*, 26(1), 5-30.
- Spiro, R. J. (1980). *Schema theory and reading comprehension: New directions* (Technical Report No. 191). Urbana, IL: University of Illinois, Center for the Study of Reading.
- Stone, R. (1985). *Effects of English/Spanish language pattern differences on ESL learners' comprehension of English text*. (ERIC Document Reproduction Service No. ED 266434)
- Tashakkori, A., & Teddlie, C. (1998). *Mixed methodology: Combining qualitative and quantitative approaches*. Thousand Oaks, CA: Sage.
- Taylor, C. & Angelis, P. (2008). The evolution of the TOEFL. In C. Chapelle, M. Enright, & J. Jamieson, (Eds.), *Building a validity argument for the Test of English as a Foreign Language* (pp. 27-54). New York, NY: Routledge.
- Tian, S. P. (2000). *TOEFL comprehension: Strategies used by Taiwanese students with coaching-school training* (Unpublished doctoral dissertation). Columbia University, New York.
- Tsui, Y. (2002). *Effects of English language ability, vocabulary knowledge, and content familiarity on comprehension performance and strategy use of high school readers learning English as a foreign language* (Unpublished doctoral dissertation). University of Kansas. Lawrence.
- Upton, T. A., & Lee-Thompson, L-C. (2001). The role of the first language in second language reading. *Studies in Second Language Acquisition*, 23(4), 469-495.
- Urquhart, S., & Weir, C. (1998). *Reading in a second language: Process, product and practice*. New York, NY: Longman.
- Valencia, S. W., & Pearson, P. D. (1987). Reading assessment: Time for a change. *The Reading Teacher*, 40(8), 726-732

- Valencia, S. W., Stallman, A. C., Commeyras, M., Pearson, P. D., & Hartman, D. K. (1991). Four measures of topical knowledge: A study of construct validity. *Reading Research Quarterly, 26*(3), 204-233.
- Van den Bergh, H. (1990). On the construct validity of multiple-choice items for reading comprehension. *Applied Psychological Measurement, 14*(1), 1-12.
- Wolf, D. F. (1993). Issues in reading comprehension assessment: Implications for the development of research instruments and classroom tests. *Foreign Language Annals, 26*(3), 322-331.
- Wu, C-P. (2005). *An investigation of metacognitive reading strategies used by EFL Taiwanese college students to comprehend familiar versus unfamiliar Chinese and English texts* (Unpublished doctoral dissertation). University of Idaho, Moscow.

APPENDIX A. STUDY INTRODUCTION FORM

Investigator: Jia-Ying Lee
668 Hawkeye Court, Iowa City, IA 52246
Tel: 319-512-9733

Advisor: Professor Lia Plakans
N274 Lindquist Hall, The University of Iowa, Iowa City, IA 52242
Tel: 319-335-5565

Description: The main purpose of this research is to compare the strategy use of Chinese-speaking students when confronted with familiar versus unfamiliar topics in a multiple-choice format reading test, such as the TOEFL iBT reading comprehension questions. The focus is on describing what students do when they are taking reading comprehension tests by asking student to verbalize their thoughts. These strategy use reports will be further examined with respect to participants' language proficiency.

Risks and Benefits: Your participation is entirely voluntary. You can choose to not participate without penalty or loss of benefits to which you are otherwise entitled. You can stop your participation at any time. There are no risks from being involved in this study. The benefits associated with this study are that you might become more aware of your reading strategies use by orally reporting your thinking processes, and become more reflective about your strategy use in the future.

Time involvement: Total time estimated to participate in this study is approximately one half to two hours.

Compensation: You will not receive any payment in this study.

Data Storage: Your identity will be kept confidential. All the data collected (participants' scores on reading assessment, think-aloud protocols) will be locked in the researcher's computer. In these cases, the data will contain no identifying information that could associate you with it, or with your participation in any study. All publications will exclude any information that will make it possible to identify you as a subject.

Contacts and Questions: If you have any questions about the study, please ask now. If you have any questions later, want more information, or wish to withdraw from the study, please call the researcher conducting the study. The names, phone numbers, and e-mail addresses are at the beginning of this form. If you have questions about your rights as a research participant, complaints, concerns, or questions about the research, you can contact

Human Subjects Office, University of Iowa
Office of the Vice President for Research
Hardin Library for the Health Sciences, Office 105

600 Newton Rd, Iowa City, IA 52242-1098
Tel: (319) 335-6564 Fax: (319) 335-7310
Email: irb@uiowa.edu

**APPENDIX B. PERMISSION LETTER FROM
HARVARD PRESS**

您好:

我是李佳盈..美國愛荷華大學博士班學生...想請問是否可以使用貴出版社之
"2007-2009 ibt 托福閱讀完全攻略"..來當論文收集資料工具書?

謝謝

敬上

李同學你好：
已問過作者，他表示沒問題，但請標示出處即可。

哈佛英語出版社

Translation:

This is to grant Jia-Ying Lee the permission to employ the following materials as the data collection instruments for her dissertation data.

Book: *Comprehensive TOEFL iBT Reading* 2007-2009

Materials: Civil Law, P.244-253
Law of Demand, P.314-323

Editor, Harvard Press (editor@howardstudy.com.tw)

Jan 10th, 2011

**APPENDIX C. PERMISSION LETTER FROM
JINNI PUBLISHING CORPORATION**

您好:

我是李佳盈..美國愛荷華大學博士班學生...想請問是否可以使用貴出版社之 " TOEFL-ibt 高分托福閱讀 120 Reading "..來當論文收集資料工具書?

謝謝

敬上

李同學你好：
已問過作者，他表示沒問題，但請標示出處即可。

知英英語出版社

Translation:

This is to grant Jia-Ying Lee the permission to employ the following materials as the data collection instruments for her dissertation data.

Book: *TOEFL iBT Reading 120*

Materials: Reflection in Teaching, pp.124-128
Electricity from Wind, pp.173-177

Editor, Jinni Publishing Corporation (toefl.cbt@msa.hinet.net)

Jan 15th, 2011

APPENDIX D. iBT PRACTICE TEST – CIVIL LAW

Question

Pause End Test

Review Help Back Next

Civil Law

The legal system is divided into civil and criminal law. Civil lawsuits have several advantages over criminal lawsuits. Civil law is defined as a body of laws regulating relations between individuals or between individuals and corporations. Issues such as property rights, personal dignity and freedom, and personal injury are protected by civil law. In some cases, legislative statutes, such as the Civil Rights Act, establish specific aspects of civil law. In others, where no particular statute exists, custom and a body of previous court decisions, collectively known as **common law**, establish precedents that constitute a working definition of individual rights and responsibilities. **Tort law** (the word 'tort' is derived from the Middle English word for injury) is a body of laws that defines compensation for damages. This kind of civil action is usually initiated by the attorney representing the injured party, also called the plaintiff. The defendant in a civil case has a right to be tried by a jury, but in highly technical issues this right often is waived and the case is heard only by the judge. Being found guilty of a civil offense can result in financial penalties, but it cannot result in jail time. In contrast to a criminal case where the burden of proof lies with the prosecution and defendants are considered innocent until proven guilty, civil cases can be decided on a preponderance of evidence. This makes civil cases considerably easier to win than criminal cases where the evidence is ambiguous. A number of mitigating factors also are taken into account in determining guilt and assigning penalties in civil cases. Culpability is based on whether the defendant could reasonably have anticipated and avoided the offense.

Reading

Question

Pause End Test

Review Help Back Next

A good-faith effort to comply or solve the problem can be a factor. The compliance history is important. ■ Is this a first offense or a habitual problem? ■ Finally, is there evidence of economic benefit to the perpetrator? ■ If so, it is more likely that willful intent was involved. ■ Most people consider being convicted of a criminal offense much more serious than losing a civil case, because the former can lead to incarceration while the latter only costs money. However, civil judgments can be extremely expensive, usually more expensive than judgments in criminal cases. A group of Alaskan fishermen won \$5 billion from the Exxon oil company for damages caused by the 1989 Exxon Valdez oil spill. Civil cases can be brought in both state and federal court. And in January 2000, the Koch Oil Company, one of the largest pipeline and refinery operators in the United States, agreed to pay \$35 million in fines and penalties to state and federal authorities for negligence in more than 300 oil spills in Texas, Oklahoma, Kansas, Alabama, Louisiana, and Missouri between 1990 and 1997. Koch also agreed to spend more than \$1 billion on clean-up and improve operations. In 2004, EarthLink, an Atlanta based Internet services company won a \$16.4 million civil suit against Howard Carmack for inundating their clients with SPAM emails.

Reading

Pause End Text

Review Help Back Next

Sometimes the purpose of a civil suit is to seek an injunction or some other form of equitable relief from the actions of an individual, a corporation, or a government agency. You might ask the courts, for example, to order the government to cease and desist from activities that are in violation of either the spirit or the letter of the law. This sort of civil action is heard only by a judge; no jury is present. Environmental groups have been very successful in asking courts to stop logging and mining operations, to enforce implementation of the endangered species act, to require agencies to enforce air and water pollution laws, and a host of other efforts to protect the environment and conserve natural resources. Often, rather than sue a corporation directly for environmental damage, it is more effective to sue the government for not enforcing laws that would have prevented the damage. A big corporation with deep pockets can afford team of lawyers and may have the resources and incentive to tie up litigation for years with motions and counter suits. Federal or state agencies may be more inclined to agree that you are right and to be willing to settle the matter quickly.

Reading

1-3

Pause End Text

Review

1. The word **others** in the passage refers to

- laws.
- decisions.
- cases.
- statutes.

2. Which of the following can be inferred about Tort cases?

- They are tried without a jury.
- They are about property rights.
- They are tried without a judge.
- They are an aspect of criminal law.

3. The word **waived** in the passage is closest in meaning to

- yielded.
- requested.
- changed.
- challenged.

(paragraph 1)

The legal system is divided into civil criminal law. Civil lawsuits have several : over criminal lawsuits. Civil law is define of laws regulating relations between indiv between individuals and corporations. Iss property rights, personal dignity and free personal injury are protected by civil law. cases, legislative statutes, such as the Civ establish specific aspects of civil law. In c no particular statute exists, custom and a previous court decisions, collectively kno **common law**, establish precedents that c working definition of individual rights an responsibilities. **Tort law** (the word 'tort' from the Middle English word for injury) laws that defines compensation for dama kind of civil action is usually initiated by representing the injured party, also called The defendant in a civil case has a right t a jury, but in highly technical issues this : waived and the case is heard only by the found guilty of a civil offense can result : penalties, but it cannot result in jail time. a criminal case where the burden of pro prosecution and defendants are considere until proven guilty, civil cases can be dec **preponderance** of evidence. This makes c considerably easier to win than criminal the evidence is ambiguous. A number of factors also are taken into account in det and assigning penalties in civil cases. Cu based on whether the defendant could re anticipated and avoided the offense.

Title	Question	Options	Navigation	Task List
Reading	4-5	Pause End Test	Back Next	

4. The word **preponderance** in the passage is closest in meaning to

- majority.
- submission.
- lack.
- demonstration.

5. According to the author, why do plaintiffs often prefer civil trials?

- It is easier to prove guilt.
- They are allowed to initiate the sentence.
- They are very technical.
- These cases avoid common law.

(paragraph 1)

The legal system is divided into civil and criminal law. Civil lawsuits have several advantages over criminal lawsuits. Civil law is defined as a body of laws regulating relations between individuals or between individuals and corporations. Issues such as property rights, personal dignity and freedom, and personal injury are protected by civil law. In some cases, legislative statutes, such as the Civil Rights Act, establish specific aspects of civil law. In others, where no particular statute exists, custom and a body of previous court decisions, collectively known as **common law**, establish precedents that constitute a working definition of individual rights and responsibilities. **Tort law** (the word 'tort' is derived from the Middle English word for injury) is a body of laws that defines compensation for damages. This kind of civil action is usually initiated by the attorney representing the injured party, also called the plaintiff. The defendant in a civil case has a right to be tried by a jury, but in highly technical issues this right often is waived and the case is heard only by the judge. Being found guilty of a civil offense can result in financial penalties, but it cannot result in jail time. In contrast to a criminal case where the burden of proof lies with the prosecution and defendants are considered innocent until proven guilty, civil cases can be decided on a preponderance of evidence. This makes civil cases considerably easier to win than criminal cases where the evidence is ambiguous. A number of mitigating factors also are taken into account in determining guilt and assigning penalties in civil cases. Culpability is based on whether the defendant could reasonably have anticipated and avoided the offense.

Title	Question	Options	Navigation	Task List
Reading	6	Pause End Test	Back Next	

(paragraph 2)

A good-faith effort to comply or solve the problem can be a factor. The compliance history is important. ■ Is this a first offense or a habitual prot

■ Finally, is there evidence of economic benefit to the perpetrator? ■ If is more likely that willful intent was involved. ■ Most people consider be convicted of a criminal offense much more serious than losing a civil case, because the former can lead to incarceration while the latter only costs mon

However, civil judgments can be extremely expensive, usually more expen than judgments in criminal cases. A group of Alaskan fishermen won \$5 bil from the Exxon oil company for damages caused by the 1989 Exxon Valde spill. Civil cases can be brought in both state and federal court. And in Janu 2000, the Koch Oil Company, one of the largest pipeline and refinery opera the United States, agreed to pay \$35million in fines and penalties to state a federal authorities for negligence in more than 300 oil spills in Texas, Okla Kansas, Alabama, Louisiana, and Missouri between 1990 and 1997. Koch : agreed to spend more than \$1 billion on clean-up and improve operations. I 2004, EarthLink, an Atlanta based Internet services company won a \$16.4 : civil suit against Howard Carmack for inundating their clients with SPAM emails.

6. Look at the four squares[■] that indicate where the following sentence coul added to the passage.

That is, did the violator gain personally from the action?

Where would the sentence best fit?

Click on a square [■]to add the sentence to the passage.

Title	Question	Option	Flag Answer	Testing Tools
Reading	7-9	Pause End Test		Review Help Back Next
7. The phrase habitual in the passage is closest in meaning to	(paragraph 2) A good-faith effort to comply or solve the problem can be a factor. The compliance history is important. ■ Is this a first offense or a habitual problem? ■ Finally, is there evidence of economic benefit to the perpetrator? ■ If so, it is more likely that willful intent was involved. ■ Most people consider being convicted of a criminal offense much more serious than losing a civil case, because the former can lead to incarceration while the latter only costs money. However, civil judgments can be extremely expensive, usually more expensive than judgments in criminal cases. A group of Alaskan fishermen won \$5 billion from the Exxon oil company for damages caused by the 1989 Exxon Valdez oil spill. Civil cases can be brought in both state and federal court. And in January 2000, the Koch Oil Company, one of the largest pipeline and refinery operators in the United States, agreed to pay \$35million in fines and penalties to state and federal authorities for negligence in more than 300 oil spills in Texas, Oklahoma, Kansas, Alabama, Louisiana, and Missouri between 1990 and 1997. Koch also agreed to spend more than \$1 billion on clean-up and improve operations. In 2004, EarthLink, an Atlanta based Internet services company won a \$16.4 million civil suit against Howard Carmack for inundating their clients with SPAM emails.			
8. The word latter refers to which of the following?		<input type="radio"/> recurring. <input type="radio"/> lively. <input type="radio"/> serious. <input type="radio"/> small.		
9. According to paragraph 2, why do defendants prefer being convicted in civil cases to being convicted in criminal cases?		<input type="radio"/> People. <input type="radio"/> Incarceration. <input type="radio"/> Criminal offenses. <input type="radio"/> Losing civil cases.		

Title	Question	Option	Flag Answer	Testing Tools
Reading	10-11	Pause End Test		Review Help Back Next
10. The author uses this phrase to enforce implementation of the endangered species act, to require agencies to enforce air and water pollution laws as two examples of	(paragraph 3) Sometimes the purpose of a civil suit is to seek an injunction or some other form of equitable relief from the actions of an individual, a corporation, or a government agency. You might ask the courts, for example, to order the government to cease and desist from activities that are in violation of either the spirit or the letter of the law. This sort of civil action is heard only by a judge; no jury is present. Environmental groups have been very successful in asking courts to stop logging and mining operations, to enforce implementation of the endangered species act, to require agencies to enforce air and water pollution laws, and a host of other efforts to protect the environment and conserve natural resources. Often, rather than sue a corporation directly for environmental damage, it is more effective to sue the government for not enforcing laws that would have prevented the damage. A big corporation with deep pockets can afford team of lawyers and may have the resources and incentive to tie up litigation for years with motions and counter suits. Federal or state agencies may be more inclined to agree that you are right and to be willing to settle the matter quickly.	<input type="radio"/> civil judgments that are expensive. <input type="radio"/> environmental groups that have changed their policies. <input type="radio"/> environmental organizations who pursued civil cases. <input type="radio"/> mining and logging organizations that have been successful.		
11. The phrase deep pockets in the passage is closest in meaning to		<input type="radio"/> a lot of money. <input type="radio"/> many employees. <input type="radio"/> a lot of inventory. <input type="radio"/> many connections.		

APPENDIX E. iBT PRACTICE TEST – LAW OF DEMAND

托福閱讀完全攻略

LAW OF DEMAND

As part of an urban revitalization project, the NY city council decided to renovate a downtown park. People from surrounding communities gathered to celebrate its reopening. Spotting a good opportunity to promote their product, the owners of a local ice cream store across from the park set up a stand, with a large sign announcing: "FREE ICE CREAM." Soon scores of people were lined up waiting to try the free ice cream. The ice cream was plentiful, and because it was free, everyone could afford it, or so it seemed. In fact, many people who wanted ice cream that day never got any. The reason of course was that they found waiting in a long line too steep a price.

When a good or service is limited, it must somehow be rationed among competing consumers. In most markets, monetary prices perform that task. But in this case of a stand offering free ice cream, waiting time becomes the effective rationing mechanism. Having to stand in line is a cost, no less a cost than spending money. This shows that although the demand curve is usually described as a relationship between the quantity demanded for a good and its monetary price, the relationship is really much more general. The demand curve is a relationship between the quantity demanded and all costs, monetary and non-monetary, associated with acquiring a good. We can thus restate the law of demand as follows. People do less of what they want to do as the cost of doing it rises.

第十組閱讀測驗 315

By stating the law of demand this way, we can see it as a direct consequence of the cost-benefit principle, which says that an activity should be pursued if its benefits are at least as great as its costs. We measure the benefit of an activity by the highest price we'd be willing to pay to pursue it. When the cost of an activity rises, it is more likely to exceed our reservation price, and we are therefore less likely to pursue that activity. The law of demand applies to BMWs, textbooks, medical care, acid-free rain and "free" ice cream. It stresses that a cost is the sum of all the sacrifices - monetary and non-monetary, implicit and explicit - we must make to engage in an activity.

In everyday language, we distinguish between goods and services people need and ones they merely want. For example, we might say that someone wants a ski vacation in Utah, but what he really needs is a few days off from work; or that someone wants a house with a view, but what she really needs is a shelter from the elements. Likewise, since people need protein to survive, we might say a severely malnourished person needs more protein. But it would strike us as odd to say that anyone, even a malnourished person, needs more steak tartar, since health can be restored by consuming far less expensive sources of protein. Economists like to emphasize that once we have achieved bare subsistence levels of consumption (the amount of food, shelter, and clothing required to maintain our health) we can abandon all reference to needs and speak only in terms of wants. This linguistic distinction helps us think more clearly about the true nature of our choices.

ESC	Question	Options	Next Section	Testing Tools
Reading		Pause End Test	Home Help Back Next	

Why does Arizona experience chronic water shortages? Some might respond that the state must serve the needs of a large population with a relatively low average annual rainfall. Yet other states, like New Mexico, also have little rainfall and do not experience water shortages nearly as often as Arizona does. Arizona's problem exists because local governments sell water at extremely low prices, which encourage Arizonians to use water in ways that make no sense for a state with low rainfall. ■ For instance oranges, which are well suited for conditions in high-rainfall states like Florida, require irrigation in Arizona. ■ But because Arizona farmers can obtain water so cheaply, they plant and flood thousands of acres of orange trees each in the Mohave Desert. ■

■ Likewise, cheap water encourages homeowners in Mesa and Scottsdale to plant water-intensive lawns and shrubs, such as the ones common in the East and Midwest. By contrast, residents of cities like Santa Fe, New Mexico, where water prices are high, choose native plants that require little watering.

Title	Question	Options	Next Section	Testing Tools
Reading	1-2	Pause End Test	Home Help Back Next	

1. The word they in paragraph 1 refers to

owners

scores

people

parks

2. The author mentions the example of the free ice-cream stand

to show that not all costs are monetary.

to indicate that short-term losses may lead to long-term gains.

to prove increasing supply lessens demand.

to describe the effects of promotions on market share.

(paragraphs 1-2)

As part of an urban revitalization project, the NY city council decided to renovate a downtown park. People from surrounding communities gathered to celebrate its reopening. Spotting a good opportunity to promote their product, the owners of a local ice cream store across from the park set up a stand, with a large sign announcing: "FREE ICE CREAM." Soon scores of people were lined up waiting to try the free ice cream. The ice cream was plentiful, and because it was free, everyone could afford it, or so it seemed. In fact, many people who wanted ice cream that day never got any. The reason of course was that they found waiting in a long line too steep a price.

When a good or service is limited, it must somehow be rationed among competing consumers. In most markets, monetary prices perform that task. But in this case of a stand offering free ice cream, waiting time becomes the effective rationing mechanism. Having to stand in line is a cost, no less a cost than spending money. This shows that although the demand curve is usually described as a relationship between the quantity demanded for a good and its monetary price, the relationship is really much more general. The demand curve is a relationship between the quantity demanded and all costs, monetary and non-monetary, associated with acquiring a good. We can thus restate the law of demand as follows. People do less of what they want to do as the cost of doing it rises.

Tab	Question	Options	Answer	Time
Reading	3-4			3:00

3. The word rationing in paragraph 2 is closest in meaning to

distributing.
 thinking.
 marketing.
 timing.

4. According to the passage, which of the following is true of the demand curve?

It is the relationship of the quantity demanded for a good or service and its monetary price.
 It is the relationship between the demand of a good or service and all costs.
 It is a general relationship.
 It is a non-monetary cost.

(paragraphs 1-2)

As part of an urban revitalization project, the NY city council decided to renovate a downtown park. People from surrounding communities gathered to celebrate its reopening. Spotting a good opportunity to promote their product, the owners of a local ice cream store across from the park set up a stand, with a large sign announcing: "FREE ICE CREAM." Soon scores of people were lined up waiting to try the free ice cream. The ice cream was plentiful, and because it was free, everyone could afford it, or so it seemed. In fact, many people who wanted ice cream that day never got any. The reason of course was that they found waiting in a long line too steep a price.

When a good or service is limited, it must somehow be rationed among competing consumers. In most markets, monetary prices perform that task. But in this case of a stand offering free ice cream, waiting time becomes the effective rationing mechanism. Having to stand in line is a cost, no less a cost than spending money. This shows that although the demand curve is usually described as a relationship between the quantity demanded for a good and its monetary price, the relationship is really much more general. The demand curve is a relationship between the quantity demanded and all costs, monetary and non-monetary, associated with acquiring a good. We can thus restate the law of demand as follows. People do less of what they want to do as the cost of doing it rises.

Tab	Question	Options	Answer	Time
Reading	5			3:00

5. Which of the sentences below best expresses the essential information in the highlighted sentence in paragraph 3? Incorrect choices change the meaning in important ways or leave out essential information.

By cost, the author means everything that is given up in order to do something.
 Every activity causes both a monetary and a non-monetary stress.
 It states that before doing any activity all costs, both explicit and implicit, must be stressed.
 All costs, regardless if they are monetary or non-monetary, are both explicit and implicit.

(paragraph 3)

By stating the law of demand this way, we can see it as a direct consequence of the cost-benefit principle, which says that an activity should be pursued if its benefits are at least as great as its costs. We measure the benefit of an activity by the highest price we'd be willing to pay to pursue it. When the cost of an activity rises, it is more likely to exceed our reservation price, and we are therefore less likely to pursue that activity. The law of demand applies to BMWs, textbooks, medical care, acid-free rain and "free" ice cream. It stresses that a cost is the sum of all the sacrifices - monetary and non-monetary, implicit and explicit - we must make to engage in an activity.

Title	Question	Option	Text
Reading	6-8	Pause End Test	Testing Tools Home Help Back Next
6. The word elements in paragraph 4 is closest in meaning to	(paragraph 4)		
<input type="radio"/> rocks. <input type="radio"/> people. <input type="radio"/> weather. <input type="radio"/> animals.			
7. The word abandon in paragraph 5 is closest in meaning to			
<input type="radio"/> to require. <input type="radio"/> to understand. <input type="radio"/> to shorten. <input type="radio"/> to discard.			
8. According to economists mentioned in paragraph 4, we only speak in terms of wants because			
<input type="radio"/> Our wants exceed our needs. <input type="radio"/> We have already fulfilled our basic necessities. <input type="radio"/> We must be able to think more clearly about the true nature of our choices. <input type="radio"/> Our language calls for it.			
			<p>In everyday language, we distinguish between goods and services people need and ones they merely want. For example, we might say that someone wants a ski vacation in Utah, but what he really needs is a few days off from work; or that someone wants a house with a view, but what she really needs is shelter from the elements. Likewise, since people need protein to survive, we might say a severely malnourished person needs more protein. But it would strike us as odd to say that anyone, even a malnourished person, needs more steak tartar, since health can be restored by consuming far less expensive sources of protein. Economists like to emphasize that once we have achieved bare subsistence levels of consumption (the amount of food, shelter, and clothing required to maintain our health) we can abandon all reference to needs and speak only in terms of wants. This linguistic distinction helps us think more clearly about the true nature of our choices.</p>

Title	Question	Option	Text
Reading	9-10	Pause End Test	Testing Tools Home Help Back Next
9. The word which in the passage refers to	(paragraphs 5-6)		
<input type="radio"/> Prices <input type="radio"/> Governments <input type="radio"/> Problems <input type="radio"/> Ways			
10. What can you infer from the final paragraph?			
<input type="radio"/> Rainfall is greater in Scottsdale than Santa Fe. <input type="radio"/> There is more water in New Mexico than the Eastern United States. <input type="radio"/> Oranges grown in New Mexico are larger than those grown in Arizona. <input type="radio"/> Water prices are higher in Santa Fe than they are in Mesa.			<p>Why does Arizona experience chronic water shortages? Some might respond that the state must serve the needs of a large population with a relatively low average annual rainfall. Yet other states, like New Mexico, also have little rainfall and do not experience water shortages nearly as often as Arizona does. Arizona's problem exists because local governments sell water at extremely low prices, which encourage Arizonians to use water in ways that make no sense for a state with low rainfall. ■ For instance oranges, which are well suited for conditions in high-rainfall states like Florida, require irrigation in Arizona. ■ But because Arizona farmers can obtain water so cheaply, they plant and flood thousands of acres of orange trees each in the Mohave Desert. ■</p> <p>■ Likewise, cheap water encourages homeowners in Mesa and Scottsdale to plant water-intensive lawns and shrubs, such as the ones common in the East and Midwest. By contrast, residents of cities like Santa Fe, New Mexico, where water prices are high, choose native plants that require little watering.</p>

Title	Question	Option	Answer	Testing Tools
Reading	11	Pause End Test		Review Help Back Next

Why does Arizona experience chronic water shortages? Some might respond that the state must serve the needs of a large population with a relatively low average annual rainfall. Yet other states, like New Mexico, also have little rainfall and do not experience water shortages nearly as often as Arizona does. Arizona's problem exists because local governments sell water at extremely low prices, which encourage Arizonians to use water in ways that make no sense for a state with low rainfall. ■ For instance oranges, which are well suited for conditions in high-rainfall states like Florida, require irrigation in Arizona. ■ But because Arizona farmers can obtain water so cheaply, they plant and flood thousands of acres of orange trees each in the Mohave Desert. ■

■ Likewise, cheap water encourages homeowners in Mesa and Scottsdale to plant water-intensive lawns and shrubs, such as the ones common in the East and Midwest. By contrast, residents of cities like Santa Fe, New Mexico, where water prices are high, choose native plants that require little watering.

Title	Question	Option	Answer	Testing Tools
Reading	11	Pause End Test		Review Help Back Next

11. Look at the four squares [■] that indicate where the following sentence could be added to the passage.

It takes 13.8 gallons of water to produce a single orange while other fruit crops require far less water.

Where would the sentence best fit?

Title	Question	Option	Answer	Testing Tools
Reading	12	Pause End Test		Review Help Back Next

12. According to the passage, which of the following is NOT true of New Mexico? (paragraph 6)

It has a low average rainfall per year.

It does not experience as many water shortages as Arizona.

Water prices are low.

They grow plants which need little water.

■ Likewise, cheap water encourages homeowners in Mesa and Scottsdale to plant water-intensive lawns and shrubs, such as the ones common in the East and Midwest. By contrast, residents of cities like Santa Fe, New Mexico, where water prices are high, choose native plants that require little watering.

APPENDIX F. TOPICAL KNOWLEDGE**VOCABULARY ASSESSMENT****(Law)**

1. tort
2. contract
3. testimony
4. prosecute
5. lawsuit
6. trial
7. jury
8. enforce
9. plaintiff

(Engineering)

1. turbine
2. wind turbine
3. megawatt
4. carbon dioxide
5. nuclear
6. hydrogen
7. solar cell
8. pumped-storage hydropower
9. electromagnetic induction

(Business)

1. ration
2. subsistence level
3. opportunity cost
4. cost-benefit principle
5. law of demand
6. reservation price
7. monetary
8. curve
9. consumption

(Teaching)

1. discipline
2. collaborative environment
3. reflective teaching
4. pedagogy
5. strategy
6. approach
7. objective (adj)
8. guidance
9. literature (review)

APPENDIX G. INTERVIEW QUESTIONS

1. Tell me how much you know about this topic, on a scale of 1-10.
2. What is your sequence of answering multiple-choice questions in the TOEFL iBT reading comprehension test?
3. What do you perceive to be the easiest and the hardest questions in the reading tests, and how did you answer these questions?
4. How would you describe your experience of approaching familiar and unfamiliar TOEFL texts and do you adopt different reading strategies?
5. How do the reading processes compare when taking a multiple-choice question test and reading in a non-testing situation?

APPENDIX H. VOLUNTARY SIGN-UP SHEET

Dear Student:

I am a doctoral student from University of Iowa, USA. I am writing to invite you to participate in a study of the strategy use of Taiwanese students who read the TOEFL iBT reading comprehension multiple-choice questions.

The purpose of this study is to gain insights into the ways that Taiwanese students answer the multiple-choice questions in the TOEFL iBT reading section. As a subject, you will need to verbally report your thinking process when you have finished the two TOEFL iBT reading materials, with a total of 20 multiple-choice questions. Then, answer the interview questions about your perception of this reading text.

The entire procedure will be conducted in a classroom located at the university. The participation will take approximately one hour.

If you are interested in joining this study, please fill the sign-up sheet and give back to your instructors. I will contact you later to schedule your time.

Sincerely,

Jia-Ying Lee

Statement of Consent:

I have read the above information and have sufficient information to make a decision about participating in this study. I consent to participate in the study.

Signature of participant

Your contact information and available time period