Reading in a Foreign Language April 2015, Volume 27, No. 1
ISSN 1539-0578 pp. 71–95

Language learner strategy by Chinese-speaking EFL readers when comprehending familiar and unfamiliar texts

Jia-Ying Lee
Tamkang University
Taiwan

Abstract

This study compares the language learner and test-taking strategies used by Chinese-speaking graduate students when confronted with familiar versus unfamiliar topics in an English multiple-choice format reading comprehension test. Thirty-six participants at a large mid-western university performed three tasks: A content knowledge vocabulary assessment, two practice reading comprehension passages from the Test of English as a Foreign Language (TOEFL) internet-based test (iBT), and a post task interview. Two stages of analysis, qualitative and quantitative, were undertaken in this study. From the qualitative results, six broad categories of strategies emerged. For the quantitative analysis, ANOVA with repeated measures was completed, and the results showed that the types of test-taking strategies adopted by Chinese-speaking students remained similar when they read passages with familiar versus unfamiliar topics. Discussions and implications related to test-taking and reading pedagogy are provided in the conclusion.

Keywords: Language learner strategies, content knowledge, multiple-choice test, test validity, familiar text

High-stakes second language reading comprehension tests designed for higher education admissions are powerful assessments for determining and distinguishing test-takers' reading achievement. Multiple-choice items are used frequently in such testing as an efficient method to elicit responses from test-takers. However, researchers have argued that test-takers' ability to select the correct answer from a list of options in multiple-choice questions fails to measure actual reading ability. For example, a score on such items might account for ability to guess well or simply use a process of elimination to narrow to an answer, which threatens test validity as neither of these processes is identified as part of a reading construct. As Bachman and Palmer (1996) argued, test validity should be judged by whether the inferences made on the basis of language tests truly reflects examinees' language ability. To make this judgment, the processes that test-takers' consciously select and their thoughts when responding to the given language tasks should be focused on. Therefore, there has been a growing recognition to gain a better understanding about how test-taking strategies are utilized in the multiple-choice reading context. Instead of reading scores, test-takers' behaviors and strategies provide insights for test validity. To investigate test-takers' processes in reading comprehension testing, this study centered on

Chinese-speaking test-takers and their strategies in English reading when they completed simulated high-stakes multiple-choice reading comprehension questions. In addition, second language readers face additional challenges when it comes to unfamiliar topics that restrict readers' content knowledge activation. The strategies, obtained by asking readers to verbalize their thoughts, were further compared with respect to the test-takers' different levels of content knowledge on the topics of the reading passages.

Review of the Literature

Content Knowledge in Reading Comprehension

Readers' background knowledge in reading has received a lot of research attention with the notion that reading comprehension depends on information from the written text and information that is retrieved from readers' background knowledge. Background knowledge refers to the information, knowledge, experience and culture that readers bring to the printed word (Krekeler, 2006; Shapiro, 2004). A number of studies have provided empirical evidence to support the notion that background knowledge can be an important factor affecting both first language (L1) and second language (L2) reading comprehension (Brantmeier, 2005; Carrell, 1987; Hudson, 1988, 2007; Krekeler, 2006; Leeser, 2003; Liu, 2011). Background knowledge is now widely recognized as a major variable in reading comprehension process since successful comprehension is achieved through the combination of text information and readers' background knowledge. In fact, several researchers have pointed out that this non-decoding variable may affect reading comprehension much more than readers' language proficiency (Alderson, 2000; Hudson, 1988; Johnson, 1982; Shapiro, 2004). Background knowledge has been shown to occupy a facilitative role in reading comprehension.

Content knowledge, one type of background knowledge, is defined as the prior and relevant knowledge stored in readers' memories that is related to the content of a text (Carrell, 1987). This content knowledge helps readers to fill in the gaps when information in the text is not explicitly stated (Alderson, 2000; Leeser, 2003; Lin, 2002; Rumelhart, 1985). In other words, readers draw inferences by using their pre-existing knowledge to decode ambiguous messages in a text. Activating the appropriate content knowledge for a text's topic is indispensable in the reading comprehension process. Evidence shows that new information is learned and remembered most when it is connected to related prior knowledge. For example, Johnson (1982) suggested that English learners recalled more information from the familiar passages and distorted more information from unfamiliar ones. Hudson (1988) also pointed out that one of the L2 reading problems lies in the lack of activating the appropriate content knowledge. With wrong content knowledge in mind, the reader will distort the text's meaning and find reading to be a difficult, even a laborious, task. In a similar study, Recht and Leslie (1988) concluded that content knowledge is an indicator of the amount and quality of information being remembered, and this knowledge is powerful enough for poor readers to compensate for their low reading abilities. Additionally, Krekeler (2006) demonstrated the important influence of content knowledge on language performance for specific academic purposes (LSAP) reading tests. However, some researchers have tried to explain why content knowledge has not always occupied a facilitative role in reading performance in L2 reading. For example, if readers cannot

correctly identify the meaning of a printed text due to their insufficient L2 proficiency, they will have difficulty in activating the correct knowledge that appropriately fits the text (Ridgway, 1997). Shapiro (2004) echoed this statement by arguing that, while correct prior knowledge assists learning, inaccurate prior knowledge can be more detrimental than having no prior knowledge at all when reading. In Shapiro's study, students who had misconceptions about a piece of information performed less well than their counterparts who had no information at all. 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, although some studies emphasized the content knowledge contribution, others have demonstrated the complicated relationship between content knowledge and reading comprehension. These contrasting views provide motivation to investigate this effect further.

Strategies in Second Language Reading Comprehension

Second language researchers have identified reading strategies employed by readers with different language proficiency levels and in various reading contexts (Anderson, 1991; Bang & Zhao, 2007; Block, 1992; Brantmeier, 2005; Erler & Finkbeiner, 2007; Karbalaei, 2010). As Karbalaei (2010) described, reading strategies reveal the ways in which readers interact with a written text, such as their thoughts and actions when evaluating and planning their reading behaviors, and how these behaviors contribute to text comprehension. Erler and Finkbeiner (2007) conceptualized comprehension strategies as "intentional actions chosen to facilitate reading at any level of processing" (p. 189). As these statements indicate, reading strategy research focuses on activities performed by the reader to build meaning from the reading material.

A number of studies have explored the strategies that L2 readers utilize to process a text (Alsheikh, 2011; Bang & Zhao, 2007; Brantmeier, 2000; Cohen & Upton, 2007). For example, Alsheikh (2011) conducted a case study to investigate the metacognitive reading strategies of three trilingual readers, who all employed more strategies in their L2 and third language (L3) than in their L1. The results showed that the number and types of strategies increased when readers encountered texts that proved to be more difficult. Bang and Zhao (2007) conducted a research study about the transfer of literacy skills across languages, with particular attention to learners' word recognition and processing skills. The authors confirmed that readers' L1 and educational background influenced their L2 reading strategies. Pedagogical cultural and educational factors have been shown to influence readers' strategy preferences.

Strategies and Readers' L1 Pedagogical Cultural Background

In addition to reader's content knowledge, 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; 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). To study the contribution of readers' L1 literacy

experiences, researchers 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. Abbott (2006) conducted research about Chinese and Arabic 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 Arabic students were more interested in comprehending English in broad concepts and used more top-down reading methods. Abbott (2006) concluded that different cultural groups used different reading strategies, which were related to readers' language and literacy backgrounds. Therefore, approaches to comprehension of L2 reading depend on readers' L1 sociocultural and educational backgrounds (Upton & Lee-Thompson, 2001).

Strategy Use in Multiple-choice Items

Studies have compared the strategies used in language assessments versus those in non-testing contexts, and have concluded that some strategies are specific to testing situations (Anderson, 1991; Cohen, 2006, 2012; Cohen & Upton, 2007; Farr, Pritchard, & Smitten, 1990; Phakiti, 2003; Shohamy, 1984). Test-taking strategies, defined by Cohen and Upton (2007), are "test-taking processes which the respondents have selected and which they are conscious of, at least to some degree" (p. 211). In responding to a reading comprehension item, Cohen and Upton (2007) further argued that three different categories of strategies were possibly drawn: reading strategies (the process related to how examinees read the passage), test-management strategies (the process of elimination), and test-wiseness strategies (the ability to achieve the correct response without understanding the text).

According to Rupp, Ferne, and Choi (2006), multiple-choice items elicit unique test-taking strategies, which are different from other types of testing formats. However, one major concern regarding multiple-choice questions is the use of test-wiseness strategies (Allan, 1992; Cohen & Upton, 2007; Hill & Larsen, 2000; Tian, 2000). Since the questions can be answered without comprehending the passage, the multiple-choice test cannot be claimed as valid. Test validity is the degree to which appropriate inferences can be drawn from the results of the measurement (Bachman & Palmer, 1996). In order to make the judgment, the thoughts of test-takers' should be considered. Therefore, test-taking strategy studies have been increasingly important in helping to construct validity by providing descriptions of how examinees reach their responses (Cohen, 2006, 2012). Instead of focusing on reading scores, there has been a call for a better understanding of test-taking strategies. With the evidence that the expected strategies are applied, a multiple-choice reading item can be claimed as valid.

The literature review brings to light some gaps associated with these studies. 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. Unfortunately, very few, if any, studies in the reading literature have examined the strategies that test-takers adopt when they are faced with familiar and unfamiliar texts, especially in high-stakes multiple-choice exams. As Mandarin Chinese-speaking participants represent the largest group of students who took the Test of English as a Foreign Language (TOEFL) internet-based test (iBT) test to gain admission

to the U.S. in 2012, this study may bridge the gaps and provide insights into reading comprehension test taking.

The research was designed to address the following questions:

1. What comprehension strategies are used by Chinese readers in the multiple-choice reading comprehension tests?
2. Are the comprehension strategies used when reading about familiar topics significantly different from those when reading unfamiliar topics?

Methods

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 and quantitative approaches provide different perspectives of information, the limitations of one approach can be compensated for by the strengths of the other method when the two methods are integrated.

In order to answer the research questions that guided this study, I chose the exploratory design, which is defined as exploring a research context with qualitative data and then measuring it with quantitative analysis (Creswell & Plano Clark, 2011). In addition, quantitative data collection and analysis were used in the beginning of the study to screen the participants. First, all individuals who responded to the study invitation were screened based on their scores on the content knowledge vocabulary assessment of four selected academic disciplines (described further in the following data collection section). Based on scores from these assessments, only participants who were considered familiar and unfamiliar readers on any two of the four disciplines continued in this study. The assessment and the criteria for participant selection will be discussed in the following data collection section. In phase two of the study, think-aloud protocols and one-on-one interviews were collected and analyzed. In the third phase, the qualitative information obtained from phase two was operationalized into numerical data and submitted for statistical analysis. Finally, the qualitative and quantitative results were combined and interpreted together.

Participants

Graduate students from Taiwan and China studying in a large mid-western university in the United States were invited to participate in this study. A total of 48 graduate students responded to the invitation. However, 36 participants remained based on the selection criteria. Of the remaining participants, 20 were female and 16 were male; they studied in four academic disciplines: business, law, language teaching, and engineering. Fifteen out of the 36 were master's students, and 21 were pursuing their doctoral degrees at the time of the study.

Data Collection Materials and Instruments

Materials. Materials used for this study were selected from two books with high-stakes testing

preparation materials, *Comprehensive TOEFL iBT Reading* 2008-2010, published by Harvard Press (2008) and the book, *TOEFL iBT Reading 120,* published by Jinni Publishing Corporation (2006) in Taiwan (see Appendix A). According to Educational Testing Service (ETS, 2003), the reading section of the TOEFL iBT measures three broad categories of reading skills: basic comprehension, inference, 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 inference questions because the reading to learn questions are different from the standard dichotomously scored multiple-choice format.

Content knowledge vocabulary assessment. In research on content knowledge, a number of measures have been used to assess readers' content knowledge of a specific topic, but I drew heavily on guidance from Huang (2010) to develop a content knowledge vocabulary assessment to evaluate participants' knowledge of the four selected academic disciplines (law, business, language teaching, and engineering). The vocabulary assessment was chosen because it has been regarded as an appropriate evaluation for readers' knowledge of a given topic (Huang, 2010; Valencia, Stallman, Commeyras, Pearson, & Hartman, 1991). The vocabulary assessment required participants to write definitions of a word from a specific content area topic. To develop this assessment, two content advisors from the represented academic disciplines were recruited to compose vocabulary lists that they regarded as appropriate for someone familiar with that content area (see Appendix B). These content advisors were all professors or doctoral students specializing in the four disciplines. It was assumed that all presented vocabulary words could be defined without reference to the text by someone who knew the topic.

Once a list was generated and finalized, content advisors helped scoring the definitions participants wrote on the content knowledge vocabulary assessments. 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. Those words answered partially were awarded a 0.5 score. The scoring results were double checked with Cronbach's alpha analysis to retain inter-rater reliability. The inter-rater reliability estimates of the four disciplines ranged from 0.82 (law) to 0.91 (business), with the overall 0.86 satisfactory result.

For every participant, the total score in each area was computed as an indicator of his/her content knowledge about a topic. In this study, participants were selected as familiar readers when they scored above 75% in one content area and as unfamiliar readers when they scored below 25% in another on the vocabulary assessments. The cut-off scores for this analysis were set by quantitative criteria with the top and bottom 25 percentile of the participants' scores defining the two groups. This determination was deemed appropriate for the exploratory study (Gelman & Park, 2008), but can be considered a potential limitation and should be carefully considered in interpreting the results of the study. Participants who scored similarly on all of the content areas were eliminated. The vocabulary assessment result for the 36 participants is presented in Table 1.

Table 1. *Results of content knowledge vocabulary assessment*

Participant's ID	Familiar Topic/ Score	Unfamiliar Topic/ Score
1	Business / 10	Language Teaching / 2
2	Language Teaching / 8	Law/ 2.5
3	Business / 7.5	Law/ 2
4	Engineering / 8	Law/ 2.5
5	Language Teaching / 9	Engineering / 1
6	Engineering / 10	Law/ 2
7	Engineering / 8	Language Teaching / 1
8	Language Teaching / 8.5	Engineering / 1.5
9	Business / 7.5	Law/ 0
10	Business / 9	Law/ 2
11	Law/ 7.5	Engineering/ 2.5
12	Language Teaching / 8.5	Law/ 2
13	Engineering / 7.5	Business / 1.5
14	Law/ 10	Engineering / 1
15	Law/ 8	Engineering / 2
16	Engineering / 8	Language Teaching / 2
17	Business / 10	Language Teaching / 2.5
18	Language Teaching / 9	Law/ 2
19	Language Teaching / 10	Business / 1
20	Language Teaching / 9.5	Business / 1.5
21	Law/ 8	Business / 1.5
22	Engineering / 8	Business / 1
23	Business / 7.5	Engineering / 2
24	Business / 7.5	Language Teaching / 1
25	Engineering / 8	Law/ 2
26	Language Teaching / 9	Engineering / 2.5
27	Law/ 8	Engineering / 1
28	Business / 7.5	Language Teaching / 0.5
29	Engineering / 8	Business / 2
30	Law/ 8	Business / 1.5
31	Business / 7.5	Engineering / 2
32	Engineering / 8	Law/ 1
33	Language Teaching / 9	Language Teaching / 1.5
34	Law/ 8.5	Engineering / 2
35	Engineering / 8	Law/ 1.5
36	Business / 7	Engineering / 2.5

Think-aloud protocols. As readers' thoughts are hidden from outside observers, oral reports from readers are necessary to gain insight into what they 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) and in test-taking situations (Anderson, 1991; Cohen & Upton, 2007; Nevo, 1989; Tian, 2000). According to Ericsson and Simon (1984), there are two ways to conduct think-aloud protocols: the introspective and retrospective. Retrospective think-aloud protocol, which requires participants 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 reading process and task intact and thus, was adopted in this study. Bowles (2010) suggested that the limitation of forgetting, which is a disadvantage of this approach, 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 right after time was up.

Interview. A semi-structured interview was conducted to clarify ambiguous information about each participant's retrospective think-aloud protocol (see Appendix C). The semi-structured interview has the advantages of allowing individuals to lead the discussion, which provides more flexibility. During the interview, participants were asked to self-report their familiarity with the reading topic, a triangulation in data collection besides the vocabulary assessment. Then the interviewer asked participants questions regarding their reading patterns, such as the order in which they answered the reading tests. Lastly, participants' experiences in approaching familiar or unfamiliar texts were explored.

Data Collection Procedure

Participants were introduced to the retrospective think-aloud protocol, then given the topical knowledge vocabulary test. After the test, participants provided times during which they were available to meet individually with the researcher for the reading test and think-aloud session. Based on participant's performance on the topical knowledge vocabulary assessment, two specific content areas, one area that was most familiar and one that was least familiar, were selected for each participant. Consequently, participants were tested on unique combinations of two reading passages corresponding to this experimental condition. The time was limited to 20 minutes for each test, as allowed by the ETS. Participants were asked to produce the think-aloud processes after they finished the first reading test. The think-aloud protocol was recorded. There was a break of 10 minutes between the think-aloud protocol and the next test. The same procedure as in the first test 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.

Data Analysis

In this study, I analyzed the data in the following phases according to the specific mixed methods purpose described earlier. In phase One, I calculated participants' scores on their content knowledge vocabulary assessments to assign the test passage and select 36 participants for the "familiar" and "unfamiliar" grouping. After the think-aloud sessions were completed, I collected and coded the think-aloud protocols and interview responses from the participants qualitatively. In phase three, I submitted the qualitative data obtained from phase two to statistical analyses. Finally, I combined the qualitative and quantitative results and interpreted them together.

Qualitative analysis. All verbal reports provided by participants in the think-aloud protocol and the interview were recorded and transcribed. Rubrics for strategies from previous research (Alsheikh, 2011; Cohen & Upton, 2007; Farr et al., 1990; Phakiti, 2003; Tian, 2000) were consulted in the initial stage of coding scheme analysis in this study. For example, Tian (2000)

utilized three categories: technical strategies, reasoning strategies, and self-adjustment strategies. Cohen and Upton (2007) listed strategies in three categories: reading strategies, test-management strategies, and test-wiseness strategies. Alsheikh (2011) grouped three broad categories of strategies from the Survey of Reading strategies (SORS), which included Global Reading strategies (GLOB), Problem Solving strategies (PROB), and Support Reading strategies (SUP). With the previous research in mind, in this study, each response from participants was reviewed as well as analyzed by the researcher and another professor who specialized in reading strategies to determine if any part of the response met the rubrics. After a list of all strategies from the think-aloud protocols had been compiled, the framework of the strategy from previous research was based on to determine if the strategies used in this study had been addressed and, if so, how they were classified. These codes of strategies were further modified to improve clarity and to obtain credibility of this measure by another detailed discussion. Finally, all the 39 strategies from the final version of qualitative analysis were grouped into six broad categories: 1. general approaches to reading the passages, 2. information by the discourse structure of the passage, 3. vocabulary/sentence-in-context approaches, 4. multiple-choice test-management strategies, 5. test-wiseness strategies, and 6. background knowledge strategies. Each category of strategy is reported in the results section with explanations and examples.

Quantitative analysis. The score for each participant's content knowledge vocabulary assessment was calculated, and used to select participants and assign readings. After data were collected, the frequency of each category of strategies reported in the think-aloud protocols, based on the previous qualitative analysis, was calculated. Strategy frequency from the interview data was also considered in frequency count. 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.

Results

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. Six broad categories of strategies were identified in this study, and an example quotation from the participant was provided if necessary.

Strategy Category 1: General approaches

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. 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.

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 that belong to the same paragraph are read together. Then the related paragraphs are read with a search for the answers.

Ex: I read all the questions related to the same paragraph first. For example, questions 1 to 3 all belonged to paragraph 1. After reading these questions, I tried to find the clues from paragraph 1. (Participant 3, think-aloud)

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

Ex: I read the question 1 and then looked for clues in the related paragraph. Then I went to the next questions (2, 3, 4…) with the same procedure. (Participant 17, interview)

4. The participant reads a portion of the passage (that might contain a potential answer) carefully.
5. The participant reads rapidly/ skims/ skips the passage if questions are not being asked.
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.

Strategy Category 2: Discourse structure

The second strategy category was identification of important information provided by the discourse structure of the passage. This category revealed that participants took advantage of English discourse structure to analyze how the text was organized. By identifying syntactic and semantic rhetorical patterns of English, participants extracted the main idea and integrated the whole passage to facilitate their understanding. Four strategies were found in this category.

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

Ex: I focused on the first few sentences and the last few sentences in each paragraph, since those sentences indicated the main points of the whole paragraph. (Participant 21, interview)

2. The participant uses knowledge of the connecting and transitional words/phrases to predict the following passage (cause/effect, compare/contrast).

Ex: The first word in this paragraph started with "however," so I knew the author is going to bring out an opposite argument and this paragraph should say something different from the previous part. (Participant 29, interview)

3. The participant uses knowledge of organizational patterns to interrelate different parts of the passage (the introduction, the supporting parts, and the conclusion part, etc.).

Ex: When I was reading, I was very aware of the framework and different parts of the passage. For example, this is the introduction with some examples, the supporting parts, then transitions with another argument, and the conclusion, etc. (Participant 12, think-aloud)

4. The participant uses knowledge of logical or transitional phrases to clarify passage organization (therefore, first of all, on the other hand, that is to say…etc.).

Ex: While I was reading, I paid special attention to those logical connecting and transitional words, such as: first of all, on the other hand, that is to say, in conclusion…etc. (Participant 34,

interview)

Strategy Category 3: Vocabulary/sentence-in-context

The third category that emerged from the think-aloud and interview 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. Six strategies were found under this category.

1. The participant looks at sentences before or after to verify the referent of a pronoun.
2. The participant infers/confirms the meanings of new words through prefixes (re= again…).

Ex: I didn't know the word "habitual," but I guessed it was 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 20, think-aloud)

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

Ex: 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 23, interview)

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

Ex: The sentence had the structure "…good or even…." So I knew that I should select the option very positive or similar to better. (Participant 14, interview)

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

Ex: Near blank 1, the sentence mentioned "take oranges for example…," but it was just the beginning of the paragraph and was not ready to introduce the information about 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 this part was related to the sentence needed to be inserted. According to their relevance, I inserted the sentence here. (Participant 16, think-aloud)

Strategy Category 4: Multiple-choice test-management strategies

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: This question mentioned "tort cases." So I highlighted the key words from the question in my mind and looked for where the related text was located. (Participant 35, think-aloud)

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

Ex: After reading the questions, I realized that the key word "worry" in the question and the word "concern" in the text corresponded with each other. (Participant 26, think-aloud)

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

Ex: This question asked why defendants preferred being convicted in civil cases than in criminal cases. So I realized that the word "incarceration" is probably the key word of this part. (Participant 14, interview)

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.

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

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: The text mentioned that teachers rarely had the time or opportunities to view their own teaching, even others' in an objective manner, so I began to paraphrase the text with the four options in order to check if they corresponded to the text. For example, for the second option, "examine thoughtfully the possible causes of events in their classrooms…." So I think this should be the right answer. (Participant 1, interview)

Strategy Category 5: Test-wiseness

The fifth category, test-wiseness, referred to participants' ability to select the correct answer without knowing the content or using the language skills that should be tested. Strategies identified in this 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. 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).

Ex: The most useful way to answer a multiple-choice question is to delete all the impossible answers first, then select the remaining one even though it is uncertain. (Participant 35, interview)

2. The participant eliminates option(s) as contradictory to paragraph/overall passage meaning.

3. The participant eliminates options as not mentioned in designated paragraph.
4. The participant selects the option because of evenly distributing answers.
5. The participant selects the longest answer.

Strategy Category 6: Background knowledge

The sixth and the final category referred to the use of background knowledge of the content area as the basis of comprehension and selection of an answer.

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

Ex: This article talks about teacher's reflection, which seems to be familiar to me as I took a class in teacher education before. (Participant 18, interview)

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

Ex: 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 36, think-aloud protocol)

Relationship Between Strategy Use and Topic Familiarity

Six strategy categories used when reading familiar versus unfamiliar texts were identified and then their corresponding means as well as standard deviations were calculated (see Table 2). Figure 1 showed a graph based on the results from Table 2. 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 2.

As shown in Table 2 and Figure 1, the similarity in strategy use with familiar versus unfamiliar texts is apparent. This is shown through a more frequent use of multiple-choice strategies (category 4) and general reading approaches (category 1), while content knowledge strategies (category 6) and discourse structural strategies (category 2) were less frequently used.

Table 2. *Means and standard deviations of familiar and unfamiliar categories of strategies*

Strategy Frequency				
Type	Strategy	N	Mean	Std Dev
FAMILIAR	1	36	2.29	0.28
	2	36	1.35	0.52
	3	36	1.60	0.25
	4	36	2.42	0.23
	5	36	1.58	0.16
	6	36	1.18	0.33
UNFAMILIAR	1	36	2.32	0.23
	2	36	1.40	0.40
	3	36	1.68	0.18
	4	36	2.48	0.25
	5	36	1.65	0.21
	6	36	1.01	0.46
TOTAL	1	36	2.31	0.25
	2	36	1.37	0.46
	3	36	1.64	0.21
	4	36	2.47	0.24
	5	36	1.60	0.18
	6	36	1.10	0.41

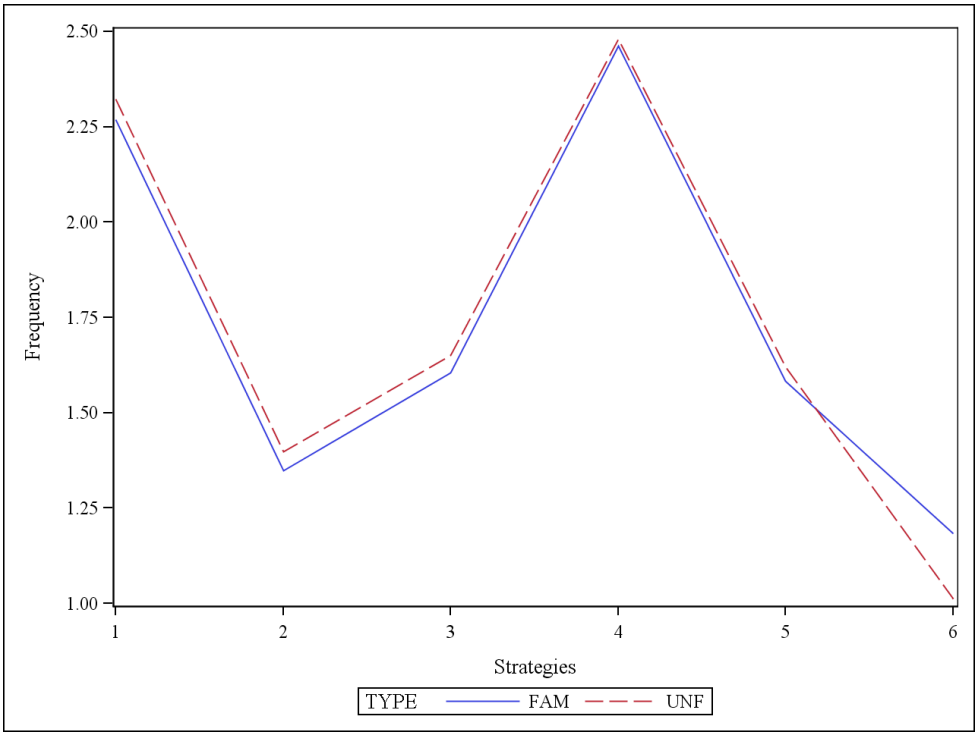

Figure 1. Distribution of strategy use

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. The results were summarized in Table 3. The 2 x 6 ANOVA analysis showed that the main effect of topic familiarity was not statistically significant (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 regardless of familiarity.

Table 3. *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	116.83	<.0001
FAM*STR	5	95	0.78	0.5532

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 4). 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 1.

Table 4. *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

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 frequently from the least and employed strategies in category 6 (content knowledge) least

frequently (see Table 2 and Figure 1).

Although the test-taking strategies used by each participant across texts with different degrees of 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. As might be expected, unfamiliar topics triggered a higher frequency of strategy use in the first five categories of this study, except for the category of content knowledge. In terms of general reading approaches, participants reread and translated certain parts of the text into Chinese more frequently to assist with their understanding of the unfamiliar text. In the category of vocabulary/sentence-in-context approaches, participants searched for more semantic and syntactic clues from the context to infer the meanings of new vocabulary for the unfamiliar texts. In multiple-choice test-management strategies, participants relied more on multiple-choice questions to identify the main points of the text and tried to paraphrase the option's meaning with text. Also, for the unfamiliar texts, participants adopted more processes of elimination and selected responses mainly because the options contained a key word that appeared in the text, which belonged to test-wiseness strategies. 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.

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 was designed to evaluate a test-takers' understanding of textual information instead of one's content knowledge, their top priority was to comprehend the text by utilizing their reading skills. For those test-takers who could not rely on their content knowledge about a topic, they felt they could still obtain the correct response when comprehending the text successfully, as expected. Thus, the participants reported that they believed content 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, participants felt that content knowledge impacted their affective responses to the testing more than it impacted their reading patterns. As expected, participants felt more confident when they had the opportunity to read a passage that was related to what they had learned in the past.

Discussion and Implications

In terms of content familiarity, participants in this study used similar strategies regardless of the texts' topic familiarity, when completing the reading comprehension test. It is possible that participants 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, Chern, & Lin, 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, participants may be accustomed to their overall test-taking techniques and employed these without making adjustments for topic familiarity, which made the reading processes that they used in comprehending both texts constant.

In addition, general or academic reading texts are designed for readers to relate the materials to their content knowledge as well as to rely appropriately on their knowledge and the text. However, the reading tests in this study, or the official TOEFL reading passages by extension, avoids building barriers based on test-takers' lack of content knowledge. High-stakes reading tests are not designed to be advantageous for some people and disadvantageous for others based on their content knowledge. This point is corroborated by ETS research reports (Hale, 1988; Liu, 2011) that the TOEFL reading comprehension passages are immune to task familiarity biases. Therefore, the texts developed to prepare test-takers for high-stakes testing, which were used in this study, are designed so that topic familiarity does not influence reading comprehension processes and performance.

Overall, participants tended to utilize more bottom-up reading strategies and encountered problems with integrating an English text as a whole (Kohn, 1992), which seemed to be influenced by Chinese pedagogical cultural background. Given that English textual structure is produced in hierarchical linguistic devices, such as superordinate references like topic sentences, the knowledge of this sequential relationship provided EFL readers with some expectations about the upcoming text structures. 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).

Among all the categories of strategies, multiple-choice test management strategies were the most frequently employed in this study. Rupp et al. (2006) argued that multiple-choice tests present a context where strategies applied are unique and are different from other testing formats. In this study, participants were aware of the advantages of multiple-choice format so the questions and options were used as indicators of which portion of the text tended to be more important or worthy of being read. That is, once participants correctly identified the related portion of the text, they devoted their attention to understand that specific part and skipped those which were not covered in the questions. Guided by the multiple-choice questions, participants were able to distinguish the main points with supplementary details. As a result, the use of multiple-choice strategies benefited participants, especially in the situations with critical time constraints.

Test-wiseness strategies and content knowledge should not be ignored in the reading curriculum, although they were not considered as primary strategies used by participants in this study. In test-taking situations, test-wiseness and content knowledge are beneficial to test-takers' performance if they are applied appropriately. Test-takers should relate important points in the text to their content knowledge in order to comprehend the text as a whole and to modify their connections according to the text content. The interview responses also supported that participants 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 (Tian, 2000).

The six categories of strategies and interview responses emerged in this study demonstrated that participants selected their answers mostly based on their understanding of the text. These results supported the validity of multiple-choice reading comprehension tests because the reading processes yielded from this given language test adequately reflected the L2 reading construct intended to be measured (Bachman & Palmer, 1996; Cohen and Upton, 2007; Daneman & Hannon, 2001). Participants in this study viewed questions and passages as one interrelated task, rather than only concentrating on the questions (Farr et al., 1990). Although the multiple-choice questions appeared to cue important parts of the texts for test-takers, 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. This substantiated 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 appeared to contradict previous studies (Cohen, 1984) in which examinees did not read or comprehend the test passages.

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. Readers from different cultures may use different reading strategies. Thus, any conclusions drawn from this study may not be generalized. This study relied heavily on participants' think-aloud protocols, a method that is limited since it cannot completely reflect readers' inner processes. The cut-off scores for the familiar versus unfamiliar groupings also limit the generalizability of the findings. This study focused only on two types of TOEFL iBT reading comprehension questions, the basic comprehension and inference questions, but 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.

To conclude, the results of this study extend the findings of Tian (2000) and Cohen and Upton (2007) by comparing the participants' test-taking strategies with different levels of topic familiarity. 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. In addition, understanding the language learner strategy related to multiple-choice questions under the simulated high-stakes reading context added insight to pedagogical implication of future reading curricula. 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*, 633–670. doi: 10.1111/j.1467-9922.2006.00391.x

Alderson, J. C. (2000). *Assessing reading.* New York, NY: Cambridge University Press.
Allan, A. (1992). Development and validation of a scale to measure test-wiseness in EFL/ESL reading test takers. *Language Testing*, *9*, 101–119. doi: 10.1177/026553229200900201
Alsheikh, N. O. (2011). Three readers, three languages, three texts: The strategic reading of multilingual and multiliterate readers. *The Reading Matrix*, *11*, 34–53.
Anderson, N. J. (1991). Individual differences in strategy use in second language reading and testing. *The Modern Language Journal*, *75*, 460–472. doi: 10.1111/j.1540-4781.1991.tb05384.x
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*, 30–50.
Block, E. (1992). See how they read: Comprehension monitoring of L1 and L2 readers. *TESOL Quarterly*, *26*, 319–341. doi: org/10.2307/3586295
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. *The Modern Language Journal*, *89*, 37–53. doi: 10.1111/j.0026-7902.2005.00264.x
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. (1987). Content and formal schemata in ESL reading. *TESOL Quarterly*, *21*, 461–481. doi: 10.2307/3586498
Cohen, A. D. (1984). On taking language tests: What the students report. *Language Testing*, *1*, 70–81. doi:10.1191/0265532202lt230oa
Cohen, A. D. (2006). The coming of age of research on test-taking strategies. *Language Assessment Quarterly*, *3*, 307–331. doi: 10.1080/15434300701333129
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*, 209–250. doi: 10.1177/0265532207076364
Cohen, A. D. (2012). Test taker strategies and task design. In G. Fulcher & F. Davidson (Eds.), *The Routledge handbook of language testing in a nutshell* (pp. 262–277). Abingdon, England: Routledge.
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*, 208–223. doi:10.1037/0096-3445.130.2.208
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.

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.

Gelman, A., & Park, D. K. (2008). Splitting a predictor at the upper quarter or third and lower quarter or third. *The American Statistician*, *62*, 1–8.

Hill, C., & Larsen, E. (2000). *Children and reading tests*. Stanford, CT: Ablex.

Hale, G., A. (1988). The interaction of student major-field group and text content in TOEFL reading comprehension. *Language Testing*, *5*, 49–61. doi:10.1002/j.2330-8516.1988.tb00257.x

Hoyt Tillman、蘇豔文〔T, H., Sue, Y.W.〕(2008). 2008-2010 iBT 托福閱讀完全攻略〔*Comprehensive TOEFL iBT Reading* 2008-2010〕. 台灣：哈佛出版社〔Taiwan: Harvard Press〕.

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*, 1–22. doi:10.1016/j.compedu.2008.06.003

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*, 503–516. doi: 10.1080/10888438.2013.774005

Karbalaei, A. (2010). A comparison of metacognitive reading strategies used by EFL and ESL readers. *The Reading Matrix*, *10*, 165–180.

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*, 99–130. doi: 10.1191/0265532206lt323oa

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*, 172–190.

Liu, Q. L. (2011). Does content knowledge affect TOEFL iBT reading performance? A confirmatory approach to differential item functioning. *Applied Measurement in Education*, *24*, 235–255. doi: 10.1002/j.2333-8504.2009.tb02186.x

Nevo, N. (1989). Test-taking strategies on a multiple-choice test of reading comprehension. *Language Testing*, *6*, 199–215.

Parry, K. (1996). Culture, literacy and L2 reading. *TESOL Quarterly*, *30*, 665–692.

Phakiti, A. (2003). A closer look at the relationship of cognitive and metacognitive strategy use to EFL reading achievement test performance. *Language Testing*, *20*, 26–56. doi: 10.1191/0265532203lt243oa

Recht, D. R., & Leslie, L. (1988). Effects of prior knowledge on good and poor readers' memory

of text. *Journal of Educational Psychology*, *80*, 16–20.
Rick Crooks、江璞〔Crooks, R., Jiang, P.〕(2006). TOEFL-iBT 高分托福閱讀 120〔TOEFL *iBT Reading* 120〕.台灣：知英文化事業有限公司〔Taiwan: Jinni Publishing Co., Ltd.〕.
Ridgway, T. (1997). Threshold of the background knowledge effect in foreign language reading. *Reading in a Foreign Language*, *11*, 151–168.
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. doi: 10.1191/0265532206lt337oa
Shapiro, A. M. (2004). How including prior knowledge as a subject variable may change outcomes of learning research. *American Educational Research Journal*, *41*, 159–189. doi: 10.3102/00028312041001159
Shohamy, E. (1984). Does the testing method make a difference? The case of reading comprehension. *Language Testing*, *1*, 147–170. doi: 10.1177/026553228400100203
Tashakkori, A., & Teddlie, C. (1998). *Mixed methodology: Combining qualitative and quantitative approaches*. Thousand Oaks, CA: Sage.
Tian, S. P. (2000). *TOEFL comprehension: Strategies used by Taiwanese students with coaching-school training* (Unpublished doctoral dissertation). Columbia University, New York.
Upton, T. A., & Lee-Thompson, L-C. (2001). The role of the first language in second language reading. *Studies in Second Language Acquisition*, *23*, 469–495.
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*, 204–233. doi: 10.2307/747761

Appendix A

iBT practice test – reflection in teaching

Teachers, it is thought, benefit from the practice of reflection, the conscious act of thinking deeply about and carefully examining the interactions and events within their own classrooms. Educators T. Wildman and J. Niles (1987) described a scheme for developing reflective practice in experienced teachers. This was **justified** by the view that reflective practice could help teachers to feel more intellectually involved in their role and work in teaching and enable them to cope with the paucity of scientific fact and the uncertainty of knowledge in the discipline of teaching.

Wildman and Niles were particularly interested in investigating the conditions under which reflection might **flourish**-a subject on which there is little guidance in the literature. They designed an experimental strategy for a group of teachers in Virginia and worked with 40 practicing teachers over several years. They were concerned that many would be "drawn to these new, refreshing conceptions of teaching only to find that the void between the abstractions and the realities of teacher reflection is too great to bridge. Reflection on a complex task such as teaching is not easy." The teachers were taken through a program of talking about teaching events, moving on to reflecting about specific issues in a supported, and later an independent, manner.

Wildman and Niles observed that systematic reflection on teaching required a sound ability to understand classroom events in an **objective** manner. They describe the initial understanding in the teachers with whom they were working as being "utilitarian...and not rich or detailed enough to drive systematic reflection." Teachers rarely have the time or opportunities to view their own or the teaching of others in an objective manner. Further observation revealed the tendency of teachers to evaluate events rather than review the contributory factors in a considered manner by, in effect, standing outside the situation.

Helping this group of teachers to revise their thinking about classroom events became central. This process took time and patience and effective trainers. The researchers estimate that the initial training of the same teachers to view events objectively took between 20 and 30 hours, with the same number of hours again being required to practice the skills of reflection.

Wildman and Niles identify three principles that facilitate reflective practice in a teaching situation. The first is support from administrators in an education system, enabling teachers to understand the requirements of reflective practice and how it relates to teaching students. The second is the availability of sufficient time and space. The teachers in the program described how they found it difficult to put aside the immediate demands of others in order to give themselves the time they needed to develop their reflective skills. The third is the development of a collaborative environment with support from other teachers. Support and encouragement were also required to help teachers in the program cope with aspects of their professional life with which they were not comfortable. Wildman and Niles make a summary comment: "Perhaps the most important thing we learned is the idea of the teacher-as-reflective-practitioner will not happen simply because it is a good or even **compelling** idea."

The work of Wildman and Niles suggests the importance of recognizing some of the difficulties of instituting reflective practice. Others have noted this, making a similar point about the teaching profession's cultural inhibitions about reflective practice. Zeichner and Liston (1987) point out the inconsistency between the role of the teacher as a (reflective) professional decision maker and the more usual role of the teacher as a technician, putting into practice the ideas of others. More basic than the cultural issues is the matter of motivation. Becoming a reflective practitioner requires extra work (Jaworski, 1993) and has only vaguely defined goals with, perhaps, little initially perceivable reward and the threat of vulnerability. Few have directly questioned what might lead a teacher to want to become reflective. Apparently, the most obvious reason for teachers to work toward reflective practice is that teacher educators think it is a good thing. There appear to be many unexplored matters about the motivation to reflect-for example, the value of externally motivated reflection as opposed to that of teachers who might reflect by habit.

1. The word justified in the passage is closest in meaning to
○ supported
○shaped
○stimulated
○suggested

2. According to paragraph 1, it was believed that reflection could help teachers
○understand intellectual principles of teaching
○strengthen their intellectual connection to their work
○use scientific fact to improve discipline and teaching
○adopt a more disciplined approach to teaching

3. The word flourish in the passage is closest in meaning to
○continue

○occur
○succeed
○apply

4. All of the following are mentioned about the experimental strategy described in paragraph 2 EXCEPT:
○It was designed so that teachers would eventually reflect without help from others
○It was used by a group of teachers over a period of years.
○It involved having teachers take part in discussions of classroom events
○It involved having teachers record in writing their reflections about teaching

5. According to paragraph 2, Wildman and Niles worried that the teachers they were working with might feel that
○the number of teachers involved in their program was too large
○the concepts of teacher reflection were so abstract that they could not be applied
○the ideas involved in reflection were actually not new and refreshing
○several years would be needed to acquire the habit of reflecting on their teaching

6. The word objective in the passage is closest in meaning to
○unbiased
○positive
○systematic
○thorough

7. According to paragraph 3, what did the teachers working with Wildman and Niles often fail to do when they attempted to practice reflection?
○Correctly calculate the amount of time needed for reflection
○Provide sufficiently detailed descriptions of the methods they used to help them reflect
○Examine thoughtfully the possible causes of events in their classrooms
○Establish realistic goals for themselves in practicing reflection

8. The word compelling in the passage is closest in meaning to
○commonly held
○persuasive
○original
○practical

9. According to paragraph 6, teachers may be discouraged from reflecting because
○it is not generally supported by teacher educators
○the benefits of reflection may not be apparent immediately
○it is impossible to teach and reflect on one's teaching at the same time
○they have often failed in their attempts to become reflective practitioners

Paragraph 4 Helping this group of teachers to revise their thinking about classroom events became central. This process took time ■ and patience and effective trainers. ■The researchers estimate that the initial training of the same teachers to view events objectively took between 20 and 30 hours, with the same number of hours again being required to practice the skills of reflection.
Paragraph 5 ■Wildman and Niles identify three principles that facilitate reflective practice in a teaching situation. ■The first is support from administrators in an education system, enabling teachers to understand the requirements of reflective practice and how it relates to teaching students.

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

However, changing teachers' thinking about reflection will not succeed unless there is support for reflection in the teaching environment.
Where could the sentence best fit?

Appendix B

Content knowledge

VOCABULARY ASSESSMENT

(Law)

1. Tort
2. Contract
3. Testimony
4. prosecute
5. lawsuit
6. trial
7. jury
8. enforce
9. plaintiff
10. court

(Engineering)

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

(Business)

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

(Teaching)

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

9. consumption	9. literature (review)
10. currency	10. principles

Appendix C

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. How would you describe your experience of approaching familiar and unfamiliar TOEFL texts and do you adopt different reading strategies?
4. How do the reading processes compare when taking a multiple-choice question test and reading in a non-testing situation?

About the Author

Jia-Ying Lee is an assistant professor of English Department at Tamkang University located in Taiwan. Her teaching and research interests focus on second language reading strategies, language assessment, and topic familiarity. E-mail: jylee@mail.tku.edu.tw