Korean EFL Students’ Strategy Use in Gap-Filling Inference Items

Yoo-Kyung Nam

(International Graduate School of English)


This paper explores Korean EFL students’ test-taking strategy use in gap-filling inference items of the College Scholastic Aptitude Test (CSAT) through retrospective think-aloud protocols. Eight college freshmen participated in this study, and verbally reported the process of answering three gap-filling inference items in their L1. Students with higher proficiency levels tended to use language learner strategies with more variety, while students of lower levels often relied on familiar words and distractors as avoidance strategies. Also, the position of the gap in each item had an effect on the order and portions of reading. Unusual reading patterns were more frequently found from those who checked the gap first, when the gap was located in the middle or at the end of the passage. To foster positive washback effect of the test items on learning, this paper suggests that teachers encourage students to effectively employ language learner strategies in order to comprehend the text rather than rely on test-wiseness.

**Key words:** CSAT, gap-filling inference item, test-taking strategies, think-aloud

1. INTRODUCTION

For recent years, Korean EFL learners have been encouraged to develop enough proficiency to read English academic texts without difficulty before they enter college (Y.-H. Park, 2010). In addition, demands for optimal item discrimination in the College Scholastic Aptitude Test (CSAT) have constantly arisen for differentiating the levels of test takers. These expectations seem to have induced a gradual increase in the length and level of reading passages in the test, consequently causing students to apply various test-taking strategies, especially to the items beyond their level. Due to the powerful washback effect

* This paper is an enhanced version of the article presented at the 2014 GETA conference.
of the CSAT on Korean high schools (Jeon, 2004), such strategies are often practiced through the “teaching to the test” approach (Bachman & Palmer, 2010) in high school classrooms. Even some after-school courses and learning materials are dedicated only to certain item types such as finding main ideas, filling in the blanks, or rearranging sentences. Despite this observable impact, it has not yet been determined through sufficient research what kind of strategies students use to answer reading comprehension items in the CSAT English subtest and whether developing those strategies would be beneficial or detrimental to their reading comprehension abilities (Oh, 1999).

While most domestic studies have employed quantitative methods to ascertain the overall patterns of Korean EFL learners’ strategy use (Choi & Chang, 2013; Hamm, 2006; Hwang, 2009; Joh, 1999; Kim & Chon, 2014; Lee, 2002, 2004; Maeng, 2006; H. Park, 2013; Y.-H. Park, 2010, 2011; H. Song, 1998; M.-J. Song, 1999), only a few studies have applied qualitative methods to investigate in depth when, how, and why those strategies were used (Jeon, 2009; Lee & Ku, 2005; Oh, 1999; Suh, 2012, 2013). In fact, more in-depth studies on these aspects are required so that teachers and item writers can have a better understanding of test takers’ strategic behavior and find alternative directions to improve reading test preparation lessons and test items in the future.

This study aims to explore Korean EFL learners’ different approaches to the CSAT English reading comprehension test by examining their strategy use in one of the most representative item types, gap-filling inference (binkan churon), where test takers are required to infer the correct word, phrase, clause, or sentence that fits in the blank in a reading passage. The gap-filling inference item holds a great significance in that it covers a noticeable proportion of the reading comprehension section, and it has been one of the most challenging items to even high level students as measured by higher order processing ability (Yamashita, 2003); some of these items in the past CSAT English subtests had the lowest percentage of correct responses for the year (J. Kim, 2014). Nevertheless, this item has been criticized by some studies that focused on strategy use in national standardized tests; not only can its answer often be inferred without discourse-level processing, but it also induces an unusual reading order in the passage due to different positions of the gaps (Oh, 1999). In addition, it has been found that gap-filling inference items elicit more strategies for correct responses based on test-wiseness (e.g., using the process of elimination) than strategies for actual reading comprehension (Kim & Chon, 2014). Focusing on this particular item type, the present study analyzes and compares Korean EFL students’ individual test-taking strategy use in depth. It also discusses implications about the influence of practicing reading with these items on L2 learning.
2. LITERATURE REVIEW

2.1. Test-Taking Strategies

Test-taking strategies are general plans that test takers use to correctly respond to test items. Bachman and Cohen (1998) referred to test-taking strategies as “test-taking processes that the respondents have selected and that they are conscious of, at least to some degree” (p. 15). In language testing, they can also be defined as language use strategies applied to assessment (Cohen, 1994).

Cohen (2012) identified three types of test-taking strategies: language learner strategies, test-management strategies, and test-wiseness strategies. Language learner strategies, not necessarily related to only tests, are methods for activating basic skills for listening, speaking, reading, and writing, including skills related to vocabulary and grammar. They “constitute the steps or actions selected by learners either to improve the learning of an L2, the use of it, or both” (Cohen, 1996, p. 5). In contrast, test-management strategies and test-wiseness strategies are directly related to test-taking situations. Test-management strategies are used for “responding meaningfully to the test item and tasks” (Cohen, 2012, p. 97) such as returning to the question for more information or comparing the options in multiple-choice items; on the other hand, test-wiseness strategies involve knowledge of the characteristics of the test format itself, such as eliminating options that seem obviously incorrect. Allan (1992) claimed that the use of test-wiseness strategies can be a strong indicator of test content/construct invalidity, sometimes revealing weaknesses in test design. He also added that once test-wiseness strategies greatly affect reading test scores, the test result would not properly account for a test taker’s reading comprehension ability.

Apart from language learner strategies, test-management and test-wiseness strategies are often investigated in studies of multiple-choice items (Cordón & Day, 1996). Despite its high reliability and practicality, the multiple-choice item has been consistently criticized for inducing test-taking strategy use only to optimize chances for choosing the right answer (Aebersold & Field, 1997; Hughes, 1989), thus not reflecting the actual reading contexts (Anderson, Bachman, Perkins & Cohen, 1991). For example, multiple-choice items often cause test takers to eliminate implausible distractors, guess answers from their general background knowledge, and focus on trivial elements that were not intended to be asked. This implies that practicing reading comprehension with multiple-choice items may hinder test takers from employing reasoning strategies that lead to successful reading comprehension (Chalmers & Walkinshaw, 2014; Hwang & Lee, 2009; Nevo, 1989), and may also facilitate the development of managing strategies to effectively answer questions within a limited time.
2.2. EFL Learners’ Strategy Use in Reading Tests

Reading comprehension test scores may characterize a learner but may not be able to explain learners’ actual thinking process toward the answer (Anderson, 1991). Thus, understanding their test-taking strategies is important for determining whether they properly comprehended the reading text. Studies of EFL learners’ reading strategies in testing situations have been conducted with different subjects, different approaches, and different foci. For domestic studies in particular, subjects have mostly been college students (Choi & Chang, 2013; H. Park, 2013; Y.-H. Park, 2010, 2011; Song, 1998; Suh, 2012, 2013), and in some other studies secondary school students as well (S.-A. Kim, 1993; Haam, 2006; Hwang, 2009). The instruments mainly used are multiple-choice items extracted from CSAT or TOEIC reading comprehension mock tests.

Quantitative studies on test-taking strategies, a large number of which have focused on the overall patterns of strategy use or the correlation between strategy use and learner variables, primarily show that reading comprehension ability (as measured by language test scores) and strategy use are significantly correlated. Some studies revealed that the higher the learner’s language proficiency is, the more the learner is aware of and uses test-taking strategies (Haam, 2006; J.-M. Kim, 2009; Kim & Chon, 2014) especially in academic contexts (Park, 2010), while others reported otherwise. Kashkouli, Barati, and Nejad Ansari (2015) found that mid-level group learners used test-taking strategies more often than high- or low-level group learners.

The variety of test-taking strategies was pointed out as another key difference between learners with higher proficiency and those with lower proficiency. Students with higher language proficiency tend to effectively use more strategies with a wider variety (e.g., using background knowledge, self-monitoring the test-taking process, inferring word meanings). They use many cognitive and metacognitive strategies consistently and meaningfully (Phakiti, 2003). In contrast, students with lower proficiency frequently use test-management strategies, mostly guessing strategies, without enough clues to infer the meaning of unfamiliar words and sentences (Hwang, 2009; Kim & Chon, 2014).

Other studies investigated the role of metacognitive awareness in using reading strategies. According to Park (2011) and Song (1999), the awareness of one’s strategy use as well as the use itself affects reading performances. Monitoring and evaluating strategy uses have been found to be particularly supportive in test takers’ reading performance (Zhang & Zhang, 2013). In addition, Choi and Chang (2013) found a significant difference in the metacognitive awareness among students of different English reading abilities: high level students showed more confidence and awareness of effective strategies in reading tasks than low level students. These findings overall seem to coincide with Grabe’s (2009) findings of reading strategy use as follows: Good readers and poor readers use the same
type of strategies, but good readers use strategies more effectively and are “more metacognitively aware of strategic responses to text difficulties” (p. 227).

On the other hand, qualitative approaches such as case studies with written reports, think-aloud protocols, interviews, and analyses have been mostly applied for exploratory purposes. Jeon (2009) conducted a case study of a Korean female student studying in the U.S. and found a discrepancy between her actual strategy use and awareness of her own strategy use. Suh (2013) conducted a similar study on four college students and found common reading strategies (setting goals, previewing, going back and forth in texts, etc.) and individual differences through cross-case analysis. Other studies have investigated how a particular item type affects test takers’ reading performances (Lee & Ku, 2005; Oh, 1999).

2.3. Gap-Filling Inference Items

Gap-filling inference is a type of inference “in which information from outside the text (general knowledge) is incorporated with information in the text to fill in gaps in missing details and to help formulate a coherent representation of the text as a whole” (Oakhill & Cain, 2007, p. 56). When combined with a multiple-choice format, a gap-filling inference item becomes an effective language testing/learning device for diagnosing learners’ inference-making ability.

In this study, a gap-filling inference item refers to a unique multiple-choice item designed for blank completion used in the CSAT and TEPS (Test of English Proficiency developed by Seoul National University). While other gap-filling items usually ask test takers to fill in short blanks with one or two words in a sentence, this type of item includes a gap, sometimes five or six words long, within a paragraph. It requires the test takers to grasp the main idea by taking either a deductive or an inductive approach: The latter refers to synthesizing the supporting details within the text, while the former refers to linking a topic sentence including a gap with another topic sentence (especially when the main idea is repeatedly mentioned in the passage) (K. Kim, 2014). Concomitantly, test takers need to find key words, use their knowledge of discourse markers, make predictions (Ahn, 2015), and understand the organization of the text such as comparison and contrast, cause and effect, or problem and solution (Jeong, 2015) in order to reach the correct answer. During this process, the use of test-management strategies (e.g., going back, rereading) and test-wiseness strategies (e.g., eliminating wrong options, choosing options including words in the passage) frequently occurs (Kim & Chon, 2014).

The advantages of this type of item are that (1) it is one of most comprehensive items for testing reading skills, such as finding the main idea, determining the logical flow of the text, and distinguishing general information from supporting details, and that (2) the scoring process is not so complicated. However, the item might ask trivial elements of the reading
passage, causing the test takers to literally translate the sentence or use limited strategies only to optimize the choices of answers. Kim and Chon (2014) voiced criticism that this type of item can be “problematic for the learners compelling them to use an array of strategies perceived to them as being successful” (pp. 81-82).

The one and only study focusing on gap-filling inference items in the CSAT is Oh’s (1999) exploratory case study using retrospective methods (written reports) with a college student. According to the study, “continuing reading” was most frequently employed, while “connecting the meaning of sentences,” “using knowledge of rhetorical structure,” and “skimming” did not occur much. The results also showed that even though the test items were originally intended for discourse-level inferencing, they could be answered correctly through sentence level processing in many cases, and that the position of gaps had an effect on altering the reading process. However, the limitations of the study are that the number of participants was only one, and that it provided limited discussion of whether the strategies used were effective for comprehending the reading passages.

Therefore, this study aims to build on the previous findings by investigating students’ test-taking strategies and discussing whether those strategies are beneficial or detrimental to reading comprehension. Below are the two research questions established for this study:

1) What are the test-taking strategies that students use to answer the items?
2) How does the reading process differ depending on the position of the gap?

3. METHOD

3.1. Participants

Eight Korean college freshmen (6 female and 2 male students) who graduated from the same high school, voluntarily participated in this study, with special interest in their strategy use in reading comprehension tests, five months after taking the CSAT in 2013.

<table>
<thead>
<tr>
<th>Student Names</th>
<th>Gender</th>
<th>Age</th>
<th>Major</th>
<th>Stanine Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alice</td>
<td>F</td>
<td>20</td>
<td>Medicine</td>
<td>1</td>
</tr>
<tr>
<td>Britney</td>
<td>F</td>
<td>20</td>
<td>Social science</td>
<td>1</td>
</tr>
<tr>
<td>Carrie</td>
<td>F</td>
<td>20</td>
<td>Business</td>
<td>2</td>
</tr>
<tr>
<td>David</td>
<td>M</td>
<td>20</td>
<td>Electric engineering</td>
<td>2</td>
</tr>
<tr>
<td>Eric</td>
<td>M</td>
<td>20</td>
<td>Physical education</td>
<td>3</td>
</tr>
<tr>
<td>Florence</td>
<td>F</td>
<td>20</td>
<td>Liberal arts</td>
<td>3</td>
</tr>
<tr>
<td>Grace</td>
<td>F</td>
<td>20</td>
<td>Nursing</td>
<td>4</td>
</tr>
<tr>
<td>Helen</td>
<td>F</td>
<td>20</td>
<td>Nursing</td>
<td>4</td>
</tr>
</tbody>
</table>
Since they had studied English focusing on CSAT preparation for more than two years, they were already familiar with the structures and item types of the test. Their average stanine scores in CSAT English subtests roughly varied from 1 to 4, and for each score two participants were selected. The basic information of the participants is shown in Table 1. In order to preserve their anonymity, their real names have been replaced by pseudonyms.

3.2. Instruments

3.2.1. Test items

Due to the possibility of having already encountered or studied items of the past CSAT exams, all the items developed by the Korea Institute of Curriculum and Evaluation (KICE) were excluded in this study to control for the familiarity of the items. Instead, items were extracted from national mock tests developed in 2014 by local district offices of education (Appendix A). As in Table 2 below, three gap-filling inference items with gaps in different positions (at the beginning, in the middle, or at the end of the passage) were selected for the main study under the assumption that test takers’ reading approaches may vary when the gaps are differently located.

<table>
<thead>
<tr>
<th>Item No.</th>
<th>Passage Topic</th>
<th>Gap Position</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Importance of prior knowledge</td>
<td>Beginning</td>
</tr>
<tr>
<td>2</td>
<td>Group identity shown in language</td>
<td>Middle</td>
</tr>
<tr>
<td>3</td>
<td>Prediction errors of the worst scenarios</td>
<td>End</td>
</tr>
</tbody>
</table>

Item 1 has a gap at the beginning of the passage, and the options are full sentences consisting of 6-8 words:

The saying that ____________ is captured in a study in which researchers wrote up a detailed description of a half inning of baseball and gave it to a group of baseball fanatics and a group of less enthusiastic fans to read.

① it takes knowledge to gain knowledge (✓)
② intelligence is much more than mere memory
③ imagination pushes the boundaries of knowledge
④ learning takes place everywhere and at all times
⑤ prejudice is an obstacle to processing information
Item 2 has a gap in the middle of the passage, and the options are polywords:

Probably the main disadvantage of speaking a mixed language involves a basic function of human language: as soon as you start to speak to someone else, your language serves as an instantly recognizable badge of your _____________.

① artistic taste ② group identity (✓) ③ intellectual power ④ personal character ⑤ educational background

Last, Item 3 has a gap at the end of the passage, and the options are verb phrases consisting of 3-4 words:

It had been built to endure the worst past historical earthquake, with the builders not imagining much worse—and not thinking that the worst past event had to be a surprise, as it _____________.

① had no precedent (✓) ② matched their prediction ③ led to better preparation ④ repeated itself over time ⑤ came earlier than expected

3.2.2. Think-aloud protocol

A think-aloud protocol is an introspective data gathering method through verbal reporting. Its strength lies in that it is the most immediate means of data elicitation (Bachman & Palmer, 2010). It is often used for investigating strategies that learners use to solve tasks. Not only does it provide teachers with rich information about strategies that should be taught in the classrooms (Hamada & Park, 2013; O’Malley & Valdez Pierce, 1996), but it also provides learners with the opportunity to become aware of their test-taking strategies (Nevo, 1989). Moreover, a think-aloud protocol allows researchers to confirm whether the participants have properly followed the directions of their studies (Morgan-Short, Heil, Botero-Moriarty & Ebert, 2012). Thus, a think-aloud protocol has a great potential in collecting useful data and supplementing quantitative methods.

Think-aloud data can be collected either retrospectively or concurrently. When subjects retrospectively think aloud, they report their thought process after carrying out a task; while when they concurrently think aloud, tasks and verbal reports are simultaneously performed. Since the instruments used in this study are reading passages with tasks requiring inferencing skills, the former was chosen as the appropriate type. While a
concurrent think-aloud protocol may cause anxiety and interfere with the test performance, a retrospective protocol not only allows participants to provide more accurate, detailed and organized reports after performing tasks (Oh, 1999), but also allows researchers to easily find whether the participants answered the item in the same way that the test constructor intended (Bachman & Cohen, 1998).

However, one of the drawbacks of this approach is that “the detail and explicitness called for by the instruction to verbalize” (Ericsson & Simon, 1993, p. 103) can negatively affect the aspects that participants focus on while they think aloud. Moreover, due to the time elapsed, a discrepancy often occurs between what was reported and what was actually thought about during the test. In order to overcome these drawbacks, the researcher asked the participants to report as immediately as possible after finishing each item, and provided some guidelines beforehand in the training session to prevent them from becoming overwhelmed by the contents they were to report.

3.3. Data Collection

3.3.1. Procedures

Since retrospective think-aloud protocol may be unfamiliar to the participants (Bowles, 2010; Cohen & Upton, 2006; Ellis & Barkhuizen, 2005; O’Malley & Valdez Pierce, 1996), instructions adapted from Green (1998) were given to the participants with a prepared script as follows:

In this study, I’m interested in how you deal with gap-filling inference items in CSAT English reading comprehension test. There will be three items, and I’ll give you three minutes for each item. Right after you answer each item, you need to tell me everything you were thinking in Korean. First, I want you to tell me about the order of reading the passage, the main idea of the text, and how you reached the answer. I may ask you follow-up questions if necessary. It’s important that you do not try too hard to give the right answer. Just be natural the way you are. This may be quite unfamiliar and awkward to you, so let me show you how to do it first and help you through a practice question.

The researcher modeled the technique in front of the participants so that they could visually understand how to report their thought process, using a different type of item (finding the main idea) from the 2014 CSAT to avoid giving hints about strategies for gap-filling inference items. Below is the model that the researcher provided:

Because this is not a long passage, I just started reading from the beginning without
skipping. In the first half of the passage, the writer talks about how we easily follow what other people want us to do. Then I paid attention when the conjunction “but” appeared in the middle. I guessed that the writer tried to oppose to what was said in the beginning. The second half was just as I expected. I finally decided the answer after reading the last two sentences. ③ seemed closest to the sentence “So don’t be afraid to say no.” [This is the basic format of your verbal reports. I might ask some questions about how you guessed the meaning of certain words or sentences. Let’s say I was asked about the word empowered. Then I would respond…] The context tells me that “feeling empowered” is a positive outcome. And because it has “power” in the word itself, I think it means that I feel like I’ve gained power.

Next, in order to create a climate similar to a real CSAT English subtest, a maximum of three minutes was given to the participants to read the texts and answer each item. Then they reported their thought processes immediately after each item using their L1 (Korean); the verbal reports were done individually rather than in pairs or groups, because “two individuals working together on a task interact, and each modifies the behavior of the other” (Green, 1998, p. 49). After each item report, the researcher asked some follow-up questions for them to explain the rationale of their strategy use in more detail if necessary.

All the verbal data were voice-recorded for later transcription. The researcher observed the participants while they were working on their tasks and took notes of any special behavior. The recorded data were transcribed into written form and then translated into English before analysis.

3.3.2. Coding scheme for analysis

The researcher designed a coding scheme by compiling existing lists of strategies. The purpose was not for marking the frequency of strategies but for maintaining intra-reliability in analysis and inferring the strategies more specifically from the verbal data (Green, 1998). Language learner strategies were based on the list of Sheorey and Mokhtari (2001), and they were assigned codes starting with LL. Test-management and test-wiseness strategies were based on the list created by Cohen (2012) and Cohen and Upton (2006), each of which was assigned codes starting with TM and TW. After coding the transcribed data based on the scheme, some strategies in the list were revised or eliminated. With the finalized scheme as in Table 3, another reviewer, currently an English teacher in a Korean high school with 5 years of teaching experience, and an MA student specializing in English education, participated in coding and cross-checking for higher inter-reliability.
### TABLE 3
Test-Taking Strategy List for Coding

<table>
<thead>
<tr>
<th>Code</th>
<th>Strategy Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>LL1</td>
<td>Using prior knowledge</td>
</tr>
<tr>
<td>LL2</td>
<td>Using context clues</td>
</tr>
<tr>
<td>LL3</td>
<td>Paraphrasing for better understanding</td>
</tr>
<tr>
<td>LL4</td>
<td>Re-reading for better understanding</td>
</tr>
<tr>
<td>LL5</td>
<td>Predicting or guessing text meanings</td>
</tr>
<tr>
<td>LL6</td>
<td>Confirming predictions</td>
</tr>
<tr>
<td>LL7</td>
<td>Noting discourse structures (cause/effect, compare/contrast, etc.)</td>
</tr>
<tr>
<td>LL8</td>
<td>Looking for markers of meaning in the passage (definitions, examples, indicators of key ideas, guides to paragraph development)</td>
</tr>
<tr>
<td>LL9</td>
<td>Guessing meaning of an unknown word, phrase or sentence</td>
</tr>
<tr>
<td>LL10</td>
<td>Looking for sentences that convey the main ideas</td>
</tr>
<tr>
<td>LL11</td>
<td>Using logical connectors to clarify content and passage organization</td>
</tr>
<tr>
<td>LL12</td>
<td>Determining what to read (often involving skipping)</td>
</tr>
<tr>
<td>LL13</td>
<td>Going back and forth in text</td>
</tr>
<tr>
<td>TM1</td>
<td>Reading the stem with the gap first and then reading the passage</td>
</tr>
<tr>
<td>TM2</td>
<td>Considering the options before going back to the passage</td>
</tr>
<tr>
<td>TM3</td>
<td>Going back to the text for clarification</td>
</tr>
<tr>
<td>TM4</td>
<td>Predicting/producing one’s own answer after reading a portion of the text</td>
</tr>
<tr>
<td>TM5</td>
<td>Identifying an option with an unknown vocabulary</td>
</tr>
<tr>
<td>TM6</td>
<td>Focusing on a familiar option</td>
</tr>
<tr>
<td>TM7</td>
<td>Selecting preliminary option(s) (lack of certainty indicated)</td>
</tr>
<tr>
<td>TM8</td>
<td>Making an educated guess (e.g., using background knowledge or extra-textual knowledge)</td>
</tr>
<tr>
<td>TW1</td>
<td>Using the process of elimination (i.e., selecting an option even though it is not understood, out a vague sense that the other options couldn’t be correct)</td>
</tr>
<tr>
<td>TW2</td>
<td>Selecting the option because it appears to have a word or phrase from the passage in it – possibly a key word</td>
</tr>
<tr>
<td>TW3</td>
<td>Guessing the correct option by combining the familiar words in the distractors</td>
</tr>
</tbody>
</table>

A few disagreements occurred in determining whether skipping was used to look for sentences that convey the main ideas or to avoid unfamiliar words and phrases; however, these were resolved through discussion based on clues in participants’ reports.

### 4. RESULT

#### 4.1. Item Response and Test-Taking Strategy Use

Through the analysis of the participants’ item responses and strategy reports, some notable features have been found in their use of test-taking strategies. For language learner strategies, the participants all reported that they carefully read the portions around discourse markers (LL8), using those markers as hints for noticing discourse structures (LL7) or distinguishing the main idea from the examples and dilatations (LL11).
Particularly, students with higher stanine scores tended to use language learner strategies with more variety. For test-wiseness strategies, “eliminating the options” (TW1) was used by every participant to some extent; some eliminated the options based on a certain rationale after reading the whole text, while the others eliminated the options with some degree of uncertainty. In addition, guessing the correct option by combining the familiar words in the distractors (TW3) was often employed mostly by students with relatively lower stanine scores to compensate for their L2 proficiency.

4.1.1. Item 1

The passage of Item 1 (Appendix A) is about an information retrieval experiment, and the gap is located at the beginning. Test takers are expected to infer the main idea (1 “It takes knowledge to gain knowledge”) after understanding the process of an experiment from the first three sentences and the results of the experiment from the rest of the passage. Five (Alice, Britney, Carrie, David, and Eric) students chose the correct answer. Below is a list of main ideas of the passage that participants stated in their verbal reports.

Alice, Eric: Those with less background information couldn’t remember important information from the story.
Britney: People who were less interested in baseball had less knowledge about it and couldn’t understand the situation, thus missing important information.
Carrie: People with less knowledge remember less information.
David: (I’m not sure about the main idea.)
Florence: If you don’t have information, you cannot remember the next one.
Grace: Environment affects runners’ performances.
Helen: You remember things well that you already know about; otherwise you forget them easily.

Among the eight participants, Alice, Britney, Carrie, Eric, and Helen’s responses clearly show that they comprehended the main idea of the passage. In the process of answering the item, Alice, Britney, Carrie and Eric chose the answer with confidence, already knowing what kind of statement would fit in the gap (TM4). They also noticed the two contrasting groups, “baseball fanatics” and “less enthusiastic fans,” and focused on their differences (LL7). In David’s case, however, even though he did correctly answer the item, he was not confident about the main idea and went through the process of eliminating the options (TW1). His example below indicates that elimination strategies may lead to the right answer but do not guarantee that he fully comprehended the passage.
David (Item 1): *(after explaining why he eliminated the other four options)* I tried to check ① one more time. Because I didn’t understand the whole text, I’m not sure if it’s the right answer. I just chose it by eliminating the wrong ones.

On the other hand, Florence misunderstood the knowledge of baseball as “prejudice” and chose the wrong answer (⑤). Grace and Helen were not able to find decisive clues and guessed the answer from the words in the distractors (TW3). Particularly Grace skimmed through the distractors and checked the content words first in order to guess the main idea before reading the passage (TM2). She did not read the whole passage but focused on finding the same or similar key words in the passage (TW2). However, such an approach led her to the wrong option.

Grace (Item 1): I got the hint from the word ‘weather.’ *(laughs)* ‘Weather’ and ‘runners advancing and runs scored.’ *(Researcher: So, you chose ⑤. What do you think the text is about?)* Umm, environment, environment affects athletes?

Although Item 1 directly asked about the main idea in the first sentence, Grace chose ⑤ (“Prejudice is an obstacle to processing information”). When asked to explain the clues that she had used, she mentioned details that were far from the main idea, focusing only on the word “obstacle” in ⑤ to match it with her guess (TM6, TW2). Thus, it can be inferred from this example that her identification of the main idea was hindered by the distractors. As for Helen, even though she had caught the gist of the passage, her lack of confidence and being affected by the distractors may have hindered her choice of the correct answer.

In addition to the process of choosing answers, the manner of dealing with unfamiliar words differed among participants. While others merely skipped unfamiliar words, Alice and Eric attempted to guess the meaning of the words (LL9).

Alice: I think I was a bit nervous that some words suddenly looked unfamiliar to me. So I tried to read here and there to understand the context (LL11) and guess the words based on the context. Now I realize that I know all the words.

Eric: I didn’t know the word ‘fanatic,’ but because the other group was ‘less enthusiastic fans,’ I guessed it might mean “enthusiastic fans.”

4.1.2. Item 2

The passage of Item 2 (Appendix A) is about how group identity is shown in language

---

1 Throughout this paper, single quotation marks indicate portions uttered in L2 (English).
development, and the gap was located in the middle of the passage. The general statements are located at the beginning and the end of the passage, with examples in the middle. All the participants except Helen answered the item correctly. As long as the passage was, many clues were embedded in the passage, and the main idea (“Your language serves as an instantly recognizable badge of your group identity”) was easier to find than the other two items without skipping. Most of their reports on the main idea of the passage were similar to some extent, mentioning language and collectivity.

Alice, Britney, David: Language shows which group you are in.
Carrie: Language reflects something collective rather than individual.
Eric: As long as the language is perfectly evolved, it doesn’t change even when people move in and out.
Grace: Languages from different regions don’t mix easily.
Helen: Language varies by region and people.

In the process of identifying the main idea, some discovered decisive clues from general statements, whereas others discovered clues from specific supporting details (LL8, LL10). Alice reported that she also utilized her experience of answering reading items in the past (TM8).

Alice: When I reached the blank, I already had a hunch that the answer would be from my background knowledge. (Researcher: What do you mean by background knowledge?) Oh, I mean past experience of taking exams. I encountered a lot of passages that had the similar main idea with this passage (TM6). But I wanted to make sure, so I read on and chose. And it was right (LL6).

Participants also utilized their knowledge of discourse markers or logical connectors to figure out the structure and the flow of the passage (LL7, LL8). For example, Alice, Britney, Eric, and Helen reported that they read the sentence starting with but carefully in order to focus on the contrast going on in the passage. Eric noticed that the word “different” repeatedly occurred in the passage and sought for clues that representing differences. He also noticed for instance and read the following example carefully to help his understanding of the general statements in the passage. Carrie and Helen paid attention to the colon in the sentence with the gap (LL11).

Carrie: Usually after a colon comes another statement repeating the same (or similar) idea to explain the previous one. So, I thought I would only read from here till the end to find some clues for the gap (LL12).
Helen: I read the sentence with the gap first (TM1), and there was a colon, which summarizes the previous part. I read it and the words ‘speak’ and ‘language’ seemed important, so I kept that in mind.

As for the vocabulary, the word “badge” was tricky for some participants since they only knew that it refers to a piece of metal or cloth but did not know it can also mean a symbol. Some skipped the word nevertheless; the others (Alice, Britney, and David) inferred the meaning as follows (LL9).

Alice: Because the text is about language representing group identity, I thought ‘badge’ is something like a sign, an indication.

Britney: Isn’t ‘badge’ some kind of a metaphor here? Just like ‘uniforms’ in the text.

David: I didn’t know what ‘badge’ meant here, so I thought for a moment, and I guessed it might be related to some kind of characteristics because it says ‘immediately recognizable.’

4.1.3. Item 3

The passage of Item 3 (Appendix A) is about people failing to prevent another negative event from happening after a previous negative event. Test takers are expected to understand what is called the Lucretius problem and apply it to the case of the tsunami that struck Japan. Only Alice chose the correct answer in this item. The difficulty results from difficult words and concepts in the passage and from the fact that the gap in this item requires the test takers to understand the example in detail. Accordingly, reports on the main idea (“not thinking that the worst past event had to be a surprise, as it ①had no precedent”) by each participant have shown large differences.

Alice: The worse can always happen than you expected. (I’m not sure if it’s right.)

Britney: Past events are not surprising, because something new always happens.

Carrie: Bad things can reoccur sometime in the future.

David: Past events are different from the present ones.

Eric: You need to be ready for worse situations.

Florence: People build something new to overcome the past, but think it would be worse. (Actually I didn’t understand what the passage was talking about.)

Grace: After recording the past earthquake, now we don’t think it was bad.

Helen: If we have some bad events in the past, something can suddenly happen.

Interestingly, none of them showed confidence in their responses this time, and their
response choices varied much more than in the previous two items. As in other items, Alice and Britney consistently read the passage from beginning to end without skipping any parts. However, Alice was able to finalize her answer (①) by eliminating options (TW1) due to not knowing the word “precedent,” while Britney failed to identify the main idea and connect her understanding of the examples with the options.

Alice (Item 3): After reading this example, I clearly understood the idea. So, I went to see the gap, I already knew what should be in the gap (TM4). But I didn’t know the word ‘precedent’ in ①. At first I was stuck, but after figuring out that others cannot be the answer, I chose ① (TW1).

Britney: In the last part, the Fukushima nuclear event? I tried to apply the first sentence to this situation. People don’t need to be surprised by past events. [pause: 5sec.] Past events aren’t necessarily surprising [voice becoming softer] because continuously, continuously, new things come, and, and they’ll be surprising as well. So, maybe ④ (“repeated itself over time”)? I don’t know what I’m talking about. [laughs]

Others focused on one or two words such as “estimate” and “historic” (TW2), attempting to match those words with the words in the distractors and eventually choosing the wrong answer.

Eric: And here, ‘information on the so called worst case scenario and use it to estimate.’ Maybe ‘estimate’ and ‘prediction’ (in ②) are related? That’s why I became sure about the answer.

Eric used this strategy only in Item 3 due to a lack of understanding of the main idea. He read the first few sentences and circled some key words. Then he found similar words from the distractors and utilized them as a tool for supporting his own interpretation, ending up in the wrong option. Another example from Grace also shows that merely matching words from the passage with the distractors might lead to inaccurate understanding, and eventually to incorrect answers.

Grace: Because there were words like ‘worst’ and ‘historic,’ I thought maybe ④ is the answer? But because of ‘not imagining much worse,’ I changed my mind and chose ③ (“led to better preparation”).

Without understanding the main idea of the passage, she chose “repeated itself over time” merely because of the words “worst” and “historic.” Then she changed her mind and
chose “led to better preparation” because of the phrase “not imagining much worse.” This may have resulted from her intention to answer the question efficiently in time. However, it has been proven through this example that the information gap cannot be filled by guessing and matching the words.

4.2. Gap Location and Reading Flow

Different approaches to the items were revealed according to the gap position in each reading passage. For example, Alice and Britney read the passage in normal order without skipping any portions regardless of the position of the gap, whereas Carrie, Eric, and Helen constantly employed the strategy of reading the sentence including the gap first (TM1). Carrie and Eric explained the reasons retrospectively as follows:

Carrie: I checked the location of the gap first. Usually if the gap is in the first sentence, it’s the main idea; the passage repeats it at the end.

Eric: Well, usually when I solve questions like this, I see the sentence with the gap first, and then look at the distractors. Umm, I learned at high school that usually the sentence with the gap is the main idea. So, I read the topic sentence, briefly catch the gist, and go on.

4.2.1. Gap located at the beginning (Item 1)

In this item, all except Helen started from the first sentence of the passage. While Alice, Britney, and Carrie identified that the passage was describing an experiment and read the passage in normal order without skipping any portions, the other five participants skipped some portions that they did not consider important (LL12). David skipped portions including detailed descriptions of the process due to a lack of understanding but instead found clues from the last sentence, while Eric skipped only the last two sentences because he already understood the main idea from the previous portions. On the other hand, Florence, Grace, and Helen failed to find the correct answer because they chose the wrong portions to skip or were distracted by other factors.

Florence: I just read the first sentence and the next one. I skipped the parts that I considered less important. (Researcher: Such as?) From the sentence starting with ‘one almost’ and the one starting with ‘the less.’ Then I focused on the reason at the end, right after the word ‘because.’
Grace: When I read the first sentence, there was a specific topic, ‘baseball.’ So, I thought it might be a part of the example. I went to the bottom, but I couldn’t find a clear main idea, so I went up again to the sentence starting with ‘afterward.’

It is inferred that Florence and Grace jumped to the end immediately after reading the first sentence because the sentences at the end were written as general statements that seem to directly convey the main idea. After having difficulty understanding the conclusion, however, Grace attempted to start with the second sentence in the passage again, while Florence did not make such effort.

Overall, when the gap was at the beginning, test takers usually started from the beginning and read the passage in normal order. However, even if they read the passage from the beginning, the portions that they selected to read or skip could be decisive in their response choices. Some tended not to use skipping strategies; instead, they followed the flow to understand the whole structure of the experiment. In contrast, the others skipped some portions either because they thought that they could understand the text without them or because they were taught to skip specific details and focus on general statements.

4.2.2. Gap located in the middle (Item 2)

Given sufficient supporting details before and after the gap, all the participants except Helen were able to choose the correct answer for Item 2. Apart from Alice and Britney, who consistently read the passage in normal order, the others’ approaches varied.

Carrie: Anyway, the gap here is slightly close to the end, so I decided not to read the first part because it’s probably just an introduction (LL12). I wanted to understand the sentence with the gap better and looked for the best place to start. Then I found ‘Why is that so?’ thinking that maybe the answer to that question would be a clue.

Eric: I started with the sentence with the gap. (Researcher: From ‘probably?’) No, actually, ‘Why is that so?’ But because it was a question I went backward to find out what the question was about and what ‘that’ refers to. Again there was ‘for instance’ /laughs/ so I went to the beginning of the passage. Eventually, I ended up starting from the beginning.

Florence: I read the options first (TM2), because they weren’t so long. And then I found the sentence with the gap and read until the end. But I didn’t understand what it was talking about /laughs/ so I went through the whole text again.

Compared with the previous case, the gap being in the middle of the passage caused
even greater differences in the amount of reading. Among those who started from the sentence including the gap, some read only the portions after the gap, while the others had to return to the beginning for a better understanding.

4.2.3. Gap located at the end (Item 3)

Alice, Britney, David, and Grace started from the first sentence without checking the gap beforehand, while Carrie and Eric, Florence, and Helen checked the gap at the last sentence first. For the latter group, the next step after checking the gap varied. Eric and Helen checked the given options first and then started with the first sentence, while Florence started with the third sentence, including the discourse marker but. However, regardless of each of their approach, relatively more portions were skipped when the gap was located at the end of the passage than at the beginning. In particular, for David and Eric, the only sentences they read were the first and the last sentences.

David: Because there is ‘past’ and ‘future risk’ in the first sentence, I thought there was going to be a ‘method’ that shows the contrast between the past and the future. So, I thought there was going to be a practical example at the end of the text, and I went straight to the last sentence.

Eric: I thought the last sentence was the topic sentence. So, I only read the first sentence and the topic sentence.

Due to the fact that the gap was located at the end, participants seem to have attempted to gather information directly related to the last sentence rather than to read the whole passage. However, as a whole, it turned out that checking the sentence with the gap first before reading the passage was not always an effective method for choosing the correct answer. Especially when the gap was located in the middle or at the end, the reading flow became unnatural. Without reading the first sentence, some of the participants did not have enough ideas to understand the rest of the passage. This caused more skipping as well; they often missed important clues and information that would help filling in the gaps.

5. CONCLUSION

This study investigated eight EFL learners’ test-taking strategies for gap-filling inference items on the CSAT through retrospective think-aloud protocols. Regarding the participants’ test-taking strategy use, some crucial points have been discovered.
According to the analysis results of the participants’ verbal reports, students with higher stanine scores on the CSAT showed more variety in their use of test-taking strategies, employing more language learner strategies such as predicting text meaning, confirming guesses, and looking for sentences that convey the main idea. Through those strategies, they were able to state the main idea more accurately and even guess what the phrase in the gap would be before reading the options. On the other hand, students with lower stanine scores used strategies mostly for guessing word meanings or deciding which portion to read and which to skip. They did actively use language learner strategies such as noticing discourse markers or logical connectors to optimize the portions to read, but whenever they experienced difficulty in finding the answer and became uncertain about their interpretation, they easily turned to test-management and test-wiseness strategies such as elimination or matching words. Even some participants only relied on one or two key words.

However, it should be noted here that having a similar stanine score on the CSAT does not necessarily imply that the test takers use similar strategies in the same situation. For example, while Carrie (stanine score 2) read the whole passage without skipping in Item 1, David (stanine score 2) looked for only general statements at first. In Item 2, Helen (stanine score 4) tried to find clues using the function of the colon within the sentence including the gap, while Grace (stanine score 4) focused on the supporting details mentioning Germany and Poland.

The position of gaps was another important factor in the reading process. While most of them read the passage from beginning to end when the gap was located at the beginning, the order and portions read started to vary greatly as they continued with the passages with gaps in the middle or at the end. When the gap was located in the middle, some even selectively read certain parts depending on linking words and punctuation after checking the gap. When the gap was located at the end, some even read only two sentences around the gap. The think-aloud data from all the three items showed that such strategic behavior (TM1) did not directly affect the rate of choosing correct answers. Yet it seems to have caused an unnatural order of reading and hindered the process of determining the main idea. Oh (1999) had pointed out this phenomenon as well, explaining that some test takers read the sentence with the gap at the end first and take the opposite direction to find clues. He also claimed that gaps need to be located at the beginning of the passage for positive washback effect and natural reading processes.

Some implications for EFL classrooms can be drawn from this study. It has been found that factors such as gap positions, difficult vocabulary or sentence structures caused the students to rely on only small portions of the passage around the gap. As a result, some could not properly understand the main idea and ended up guessing the answer based on fragmentary information. Therefore, alternative strategy training methods are needed for
test-focused classrooms to improve students’ discourse-level inferencing skills than training them to deal with items merely the way they are given. One solution is to eliminate all the distractors from the item and provide students the passage with the gap. Then, as Brown (2004) suggested, students can complete the sentences by producing a phrase. Another solution is to have students share and discuss their strategies item by item focusing on what worked and what did not. They need to be exposed to various types of strategies and ascertain for themselves which strategies would be more effective in a certain situation. Students with low language proficiency are especially likely to benefit from this approach since they need to learn to employ test-taking strategies selectively and effectively (Kim & Chon, 2014). Also, they can reflect on their strategy use by writing strategy learning logs and receiving teacher feedback (Brown, 2004), or they can enhance their metacognitive awareness by often thinking and verbalizing their strategies (Hamada & Park, 2013). However, most importantly, strategy training needs to be accompanied by lessons focused on improving linguistic competence and investing more time practicing language learner strategies than test-management or test-wiseness strategies.

Despite the effort to avoid drawbacks, the study is not without its limitations. First, the results may be difficult to generalize due to the small number of participants. Another possible weakness lies in the limitations of retrospective think-aloud protocols. Since the researcher had to assure that they were reporting their strategies as were instructed, it was inevitable for the researcher to sit in front of the participants and observe their performance. Therefore, observer’s paradox (the act of observation altering the nature of the language use) might have occurred (Bachman & Cohen, 1998). Also, each student may have had different abilities to verbalize their thinking process (Joh, 2013); a lack of report cannot be interpreted as a lack of awareness (Allport, 1988, cited in Bowles, 2010). Most importantly, the difficulty of items differed in this study: one item yielded a high correct response rate while the others did not. Thus, it is unclear whether gap position was the only factor that altered the reading process or affected strategy use. In further studies, item difficulty need to be controlled based on piloting or a reliable source of correct response rate.

Nevertheless, the results of this study imply that strategy instruction for multiple-choice reading tests needs to be delivered prudently in EFL classrooms. They need to focus more on language ability itself rather than test-wiseness features. More studies employing think-aloud methods are recommended for diagnosing Korean EFL students’ strengths and weaknesses in language learning, understanding their behavior patterns in reading tests for each item type, and evaluating the construct validity of items for improvement.
REFERENCES


APPENDIX A

Items Used for Think-Aloud Protocols

Item for the training session

In our efforts to be the good child, the uncomplaining employee, or the cooperative patient, many of us fall into the trap of trying to please people by going along with whatever they want us to do. At times, we lose track of our own boundaries and needs, and the cost of this could be our life, both symbolically and literally. When we are unable to set healthy limits, it causes distress in our relationships. But when we learn to say no to what we don’t feel like doing in order to say yes to our true self, we feel empowered, and our relationships with others improve. So don’t be afraid to say no. Try to catch yourself in the moment and use your true voice to say what you really want to say.

① 난관을 극복할 때 성취감이 생긴다.
② 항상 타인의 입장을 먼저 고려해야 한다.
Item 1 The saying that ________________ is captured in a study in which researchers wrote up a detailed description of a half inning of baseball and gave it to a group of baseball fanatics and a group of less enthusiastic fans to read. Afterward they tested how well their subjects could recall the half inning. The baseball fanatics structured their recollections around important game related events, like runners advancing and runs scored. One almost got the impression they were reading off an internal scorecard. The less enthusiastic fans remembered fewer important facts about the game and were more likely to recount superficial details like the weather. Because they lacked a detail internal representation of the game, they couldn’t process the information they were taking in. They didn’t know what was important and what was trivial. They couldn’t know what mattered. Without a conceptual framework in which to embed what they were learning, they were extremely forgetful.

1. it takes knowledge to gain knowledge
2. intelligence is much more than mere memory
3. imagination pushes the boundaries of knowledge
4. learning takes place everywhere and at all times
5. prejudice is an obstacle to processing information

Item 2 Languages evolve differences because different groups of people independently develop different words and different pronunciations over the course of time. But the question remains why those diverged languages don’t merge again when formerly separated people spread out and reconnect each other at speech boundaries. For instance, at the modern boundary between Germany and Poland, there are Polish villages near German villages, but the villagers still speak a local variety of either German or of Polish, rather than a mix of German and Polish. Why is that so? Probably the main disadvantage of speaking a mixed language involves a basic function of human language: as soon as you start to speak to someone else, your language serves as an instantly recognizable badge _________________. It’s much easier for wartime spies to wear the enemy’s uniform than to imitate convincingly the enemy’s language and pronunciation. People who speak your language are your people, whereas someone speaking a different language is apt to be regarded as a potentially dangerous stranger.
Item 3 Risk management professionals look in the past for information on the so called worst case scenario and use it to estimate future risks—this method is called “stress testing.” They take the worst historical recession, the worst war, or the worst point in unemployment as an exact estimate for the worst future outcome. But they never notice the following inconsistency: this so called worst case event, when it happened, exceeded the worst case at the time. I have called this mental defect the Lucretius problem, after the Latin poetic philosopher who wrote that the fool believes that the tallest mountain in the world will be equal to the tallest one he has observed. The same can be seen in the Fukushima nuclear reactor, which experienced a huge failure in 2011 when a tsunami struck. It had been built to endure the worst past historical earthquake, with the builders not imagining much worse—and not thinking that the worst past event had to be a surprise, as it __________.

1. had no precedent
2. matched their prediction
3. led to better preparation
4. repeated itself over time
5. came earlier than expected

Applicable levels: Tertiary

Yoo-Kyung Nam
Dept. of ELT Materials Development
International Graduate School of English
17, Yangjae-daero 81-gil, Gangdong-gu
Seoul 134-847, Korea
Phone: 02-6477-5114
Email: curlywoong@igse.ac.kr

Received on September 1, 2015
Reviewed on October 15, 2015
Revised version received on November 15, 2015