

Research Forum

The Role of Think-Aloud and Metacognitive Strategies in L2 Meaning-Inference During Reading

Megumi Hamada
Ball State University, U.S.A

Chaehee Park
Sunmoon University, Republic of Korea

Lexical inference is an important word learning method, yet it is still unknown what kind of instruction will improve inference accuracy. In this study we investigated whether think-aloud can enhance metacognitive strategies while inferring the meaning of unknown words encountered during reading and lead to higher inference accuracy. Two groups of college-level Korean L1 ESL students inferred the meanings of pseudowords in a short passage in a pretest-posttest design. The think-aloud group ($n = 19$) performed the task verbalizing their thoughts, while the control group ($n = 20$) performed it silently. The results indicated a significant gain in inference accuracy for the think-aloud group. Based on further analyses of the think-aloud protocol, the study addresses the importance of the quality of think-aloud in L2 performance.

字句推論は重要な語彙学習方法であるが、推論の正確さを向上する方法はまだ分かっていない。本研究は、読書中に出てくる未習の語彙の意味を推測する際に、声に出して考えるという方法が、メタ認知を促進することによって、高い正確率をもたらすかどうかという調査を行った。事前・事後テストのデザインで、2グループの韓国語を母語とするESL学習者が、短い文章に含まれている偽単語の意味を推測した。声に出して考えるグループ ($n = 19$) はそのタスクの間考えを声に出しながら、一方コントロールグループ ($n = 20$) はタスクを無言で行った。結果は声に出して考えるグループの推測正確度が大幅に向上したことを示した。プロトコロの更なる分析をふまえて、本研究は声に出して考える内容の質が重要であることを述べる。

Word knowledge is crucial in all aspects of second-language (L2) learning throughout the proficiency levels. In the past decade, there has been increasing interest in the nature of this knowledge and its acquisition. One growing area of research focuses on lexical inference while reading. Lexical inference, hereinafter called meaning-inference, involves “making informed guesses as to the meaning of a word in light of all available linguistic cues in combination with the learner’s general knowledge of the world and awareness of context” (Haastrup, 1991, p. 40). Meaning-inference is also known as “incidental” word learning because it can occur as a by-product of other activities, such as reading (Huckin & Coady, 1999). The indirect nature of meaning-inference is appealing, and this method is suggested to be “the most important of all sources of vocabulary learning” (Nation, 2001, p. 232).

Despite the fact that meaning-inference has been widely incorporated into teaching, teachers often encounter “wild guesses” – the fact that not every student is able to infer the correct meaning of unknown words (e.g., Kaivanpanah & Alavi, 2008; Kelly, 1990). Although a number of findings have been reported on meaning-inference, findings concerning instructional techniques or strategies that enhance inference accuracy are still limited. This study focused on the use of think-aloud during meaning-inference, based on research in reading comprehension. A think-aloud procedure, originally developed in the field of psychology, is a commonly used research technique both in first language (L1) and L2 (Bowles, 2010). This technique requires research participants to verbalize their thought processes during or after a given task, and the verbal protocol generated from the think-aloud is used to reveal the participants’ thought processes while performing the task (Ericsson & Simon, 1993; Pressley & Afflerbach, 1995). A number of L1 studies suggest that think-aloud can be used to promote the use of metacognitive strategies, which in turn facilitates reading comprehension (e.g., Baumann, Seifert-Kessell, & Jones, 1992; Loxterman, Beck, & McKeown, 1994).

With the assumption that both reading comprehension and meaning-inference during reading are outcomes of cognitive processes and facilitated by metacognitive strategies, this study focused on the relationship between think-aloud, metacognitive strategies, reading comprehension, and meaning-inference during reading in the L2 context with Korean L1 college students enrolled in an intensive English program in the United States. This college-level learner population, both in South Korea and in Japan, often has basic skills in English yet has difficulty in reading advanced or authentic materials due to a lack of vocabulary knowledge. This study aimed to provide further

findings on vocabulary learning among Northeast Asian college-level learners. To clarify the relationships between meaning-inference during reading, reading comprehension, and think-aloud, the following section presents a review of relevant research.

Literature Review

Think-Aloud

Think-aloud is categorized into two types, concurrent or retrospective, depending on the timing of the task administration. Concurrent think-aloud requires learners to verbalize their thought processes as they work on a given task, and retrospective think-aloud requires verbalization after task completion. Both types of think-aloud are intended to reveal what thought processes participants are actually going through while performing a learning activity (Ericsson & Simon, 1993; Pressley & Afflerbach, 1995). The verbal reports generated through think-aloud are further categorized as nonmetalinguistic or metalinguistic, depending on the quality of thought processes reflected in the report (Ericsson & Simon, 1993). Nonmetalinguistic (or nonmetacognitive) think-aloud is defined as a report including thoughts *per se*, while metalinguistic (or metacognitive) think-aloud is defined as including deliberate thoughts such as reasons, explanations, and justifications of specific decisions made during task performance (Bowles & Leow, 2005).

Although think-aloud has provided valuable contributions to language research, its use may be questionable due to the reactivity issue. The reactivity of think-aloud refers to the problem that concurrent think-aloud potentially alters outcome results by influencing participants' cognitive processing while they are completing a given task (Ellis, 2001; Jourdenais, 2001). The main issues investigated in this line of research are reactivity in the accuracy and latency of participants' task performance. Think-aloud may influence the accuracy on outcome measures and the latency of time on task, yet current findings are still inconclusive. For example, regarding accuracy, Sanz, Lin, Lado, Bowden, and Stafford (2009) and Rossomondo (2007) reported positive reactivity, suggesting that think-aloud led to higher gain in receptive and productive knowledge of L2 grammar introduced through written input, whereas Egi (2008) reported nonreactivity in the recall of new L2 grammar introduced through spoken input. Examining reactivity in L2 writing, Sachs and Polio (2007) reported that think-aloud is negatively reactive, suggesting that think-aloud led to a decrease in the number of errors corrected in the

revised essay in Experiment 1 (within-participant design), while reporting nonreactivity in Experiment 2 (between-participant design) on the same issue. Regarding time, some studies report negative reactivity, suggesting that think-aloud increased time on task (e.g., Bowles, 2008; Bowles & Leow, 2005; Yoshida, 2008, Experiment 1 in Sanz et al., 2009), whereas others report nonreactivity (e.g., Sachs & Suh, 2007; Experiment 2 in Sanz et al., 2009).

The reactivity of think-aloud has been investigated in reading comprehension tasks as well. A majority of findings report nonreactivity (Bowles & Leow, 2005; Leow & Morgan-Short, 2004; Rossomondo, 2007; Yoshida, 2008), although there exists research that reports otherwise (Goo, 2010). For example, in Leow and Morgan-Short (2004), 77 college-level learners of Spanish, divided into a think-aloud group and a non-think-aloud group, read Spanish texts and answered multiple-choice reading comprehension questions written in their L1, English. The think-aloud group reported to a laboratory and was instructed to voice their thoughts aloud throughout the text reading and comprehension questions, whereas the non-think-aloud group remained in the classroom and completed the tasks silently. The results indicated that there was no significant difference between the two groups' reading comprehension scores.

Likewise, Bowles and Leow (2005) examined reactivity in reading comprehension among college-level Spanish L2 students and found a nonsignificant difference between the think-aloud and the silent control groups. They also took the important step of examining the effect of type of think-aloud in reactivity. The think-aloud group was divided into a nonmetalinguistic (or nonmetacognitive) think-aloud group, in which the participants were required to say aloud whatever passed through their minds while performing the task, and a metalinguistic (or metacognitive) think-aloud group, in which they were required to say aloud more specific, additional information, such as justification and reasoning for their answers. The results indicated that the nonmetacognitive think-aloud group scored significantly better than their counterpart, implying that metacognitive think-aloud was cognitively more demanding and interfered with reading comprehension. This finding suggests that the complexity of information verbalized during think-aloud clearly plays a role in determining the existence of reactivity.

Meaning-Inference During Reading

Meaning-inference is a complex construct because it entails multiple processes such as analyzing, extracting, and integrating textual information with the reader's background knowledge. Sternberg (1987) identified the major sequential operations involved in meaning-inference during reading: (a) separating relevant from irrelevant text information for the purpose of inferring the meaning of an unknown word, (b) combining relevant textual cues to formulate a workable definition, and (c) evaluating the hypothesized meaning against information from the subsequent context. Because the main source of information or cues used for meaning-inference is the text, accurate comprehension of the text is crucial for accurate meaning-inference.

A number of L2 meaning-inference studies have investigated what cues learners tend to use in meaning-inference during reading. The cues are generally categorized as either local or global cues. Local cues include morphological structure, word analogy, and grammatical (syntactic) structure, wherein learners do not necessarily have to apply their understanding of the text in meaning-inference. For example, Huckin and Bloch (1993) investigated college-level ESL learners' meaning-inference behaviors and reported that the cue that was most important for accurate meaning-inference was the morphological structure of an unknown word.

In contrast, global cues involve more contextually based analysis that requires a deeper level of understanding of the text as well as world knowledge, such as background knowledge and the existence of schemata related to the text, rather than an analysis solely of the given text. It is important to note that learners may use multiple cues, including both local and global cues, in inferring the meaning of a single word. There are a handful of findings that suggest that using local cues alone tends to yield inaccurate inference (e.g., Chern, 1993; Haynes, 1993; Nassaji, 2003). For example, Haynes (1993) reported that an adult ESL student inferred "the end of spring" for *offspring* based on an incorrect morphological analysis (*off + spring*), despite the fact that the inferred meaning did not match the context. Similarly, Nassaji (2003) found that adult ESL learners incorrectly inferred meanings of unknown words based on graphic similarity of words (word analogy), such as the similarity of *permeated* to *meat* and *affluence* to *influence* (p. 653). Although local cues may not always lead to accurate inference, they are more popular than global cues among learners (e.g., Bensoussan & Laufer, 1984; Haynes, 1993). These findings suggest that local cues are more easily used by learners, presumably because local cues do not require as much comprehension of the text as global cues do.

Metacognition in Reading and Meaning-Inference

Metacognition is defined as “one’s knowledge concerning one’s own cognitive processes and products or anything related to them” (Flavell, 1976, p. 232). Metacognitive strategies promote conscious awareness of one’s cognitive processes, such as planning for learning, thinking about the learning process as it is taking place, monitoring of one’s production or comprehension, and evaluating learning after a task is completed (Purpura, 1997). In reading comprehension, different levels of cognitive processes are involved, such as visual, phonological, and semantic processes, word recognition, syntactic parsing, and discourse processes, and all of the processes contribute to the accurate comprehension of a text (e.g., Koda, 2005). Metacognition plays an important role in controlling these processes, and there has been a large volume of research, particularly in L1 reading instruction, that investigates metacognitive strategies that enhance reading comprehension. One strategy with empirical support for its effectiveness is think-aloud. Because think-aloud can reveal certain cognitive processes, using think-aloud while comprehending a text makes it possible to teach what cognitive processes need to be involved in reading comprehension. This point is discussed more in the next section.

In meaning-inference during reading, separating and combining textual cues and evaluating an inferred meaning (Sternberg, 1987) are involved in addition to the processes required in reading comprehension. Thus, as in reading comprehension, metacognition is necessary to facilitate the meaning-inference processes. For instance, in Nassaji (2003), 21 intermediate-level adult students with various L1 backgrounds (Arabic, Chinese, Persian, Portuguese, and Spanish), who were enrolled in a 12-week intensive English program, engaged in a meaning-inference task. Each student individually met with the researcher and was asked to read aloud a short passage which contained target words unfamiliar to them. They were then asked to infer the meaning of the words, using the think-aloud technique, in English. They were also asked to underline any other unfamiliar words in the passage and infer the meaning of the words, using think-aloud. The analysis of the verbal protocols indicated that “verifying” (evaluating whether the inferred meaning is correct) and “self-inquiry” (asking questions about a word or the inferred meaning of a word) in the think-aloud protocol were related to higher inference accuracy (p. 662), suggesting the importance of metacognitive strategies in meaning-inference during reading.

Think-Aloud in Reading Instruction

Think-aloud has also been an important technique for instructional purposes. The most prominent implementation of think-aloud is seen in reading comprehension instruction, mainly targeting L1 readers. This instructional implementation of think-aloud is based on the view that reading comprehension is a complex cognitive activity that requires a strategic planning and problem solving process (Kucan & Beck, 1997). Because think-aloud is able to reveal learners' cognitive processes, using think-aloud while comprehending a text makes it possible to teach what cognitive processes need to be involved for reading comprehension. In other words, what is essentially taught using think-aloud are metacognitive strategies that facilitate comprehension processes.

In order to incorporate metacognitive strategies, typical think-aloud instruction includes teacher modeling, using concurrent think-aloud, to show the way comprehension should be carried out and discussion questions or activities that direct students to be engaged in their comprehension processes (e.g., Maria & Hathaway, 1993; Oster, 2001; Walker, 2005). The effectiveness of think-aloud instruction in L1 reading comprehension has been reported in many studies (e.g., Baumann et al., 1992; Bereiter & Bird, 1985; Berne, 2004; Laing & Kamhi, 2002). For example, in Laing and Kamhi (2002), 40 third-graders were divided into groups of average and below-average readers and instructed with two conditions: listen through and think-aloud. After reading a story, the children answered comprehension questions. As expected, the comprehension accuracy was significantly better for both groups in the think-aloud condition. Given the empirical support, think-aloud instruction has practically become a standard method in reading comprehension instruction with native speaking students.

As for the use of think-aloud in L2 reading comprehension instruction, current findings are far fewer than in L1 research, despite the fact that there have been a number of suggestions and recommendations that metacognitive strategies are crucial in L2 reading comprehension (e.g., Block, 1992; Carrell, 1989; Casanave, 1988). As for empirical studies, McKeown and Gentilucci (2007) investigated whether think-aloud instruction, using concurrent think-aloud, improves middle school ESL learners' content area reading comprehension. Due to a small number of participants ($N = 27$), all of them did a pretest, treatment (think-aloud instruction), and a posttest. In contrast to findings from L1 studies, the results indicated that none of the student groups (early intermediate, intermediate, and early advanced) showed a reliable gain in reading comprehension scores in the posttest. However, two

issues need to be noted in interpreting their results. There was no control group, which makes the research design of this study questionable. Another issue is that the think-aloud instructions were administered in English and the learners were asked to think-aloud in English, rather than their L1. Given the fact that the students had relatively low English proficiency, it is possible that think-aloud in their L2 did not completely reflect their metacognition.

Research Question

As summarized above, in reading instruction, think-aloud is widely used as an activity or technique to enhance the use of metacognitive strategies, that is, to make students aware of their own cognitive processes in reading comprehension. The underlying assumption is that think-aloud can reveal certain cognitive processes (Ericsson & Simon, 1993). In other words, being able to think-aloud their own cognitive processes indicates that the learners are aware of the actual processes. Cognitive processes are involved in not only reading comprehension, but in meaning-inference as well. That is, if metacognitive strategies play a role in reading comprehension, it seems reasonable to argue that metacognitive strategies also play a role in meaning-inference, because both activities involve cognitive processes. Although a number of studies in L2 meaning-inference during reading have used think-aloud as a research technique for analyzing learners' cognitive processes (e.g., Nassaji, 2003; Paribakht & Wesche, 1999), whether the use of think-aloud influences the outcome of meaning-inference has not been extensively examined. Thus, the research question investigated in this study was:

Does the use of think-aloud while inferring the meaning of unknown words encountered during reading improve inference accuracy?

The research question was tested using a pretest-posttest design with two groups of participants, a control group vs. an experimental group. The two groups varied in terms of treatment – whether or not think-aloud was included in the treatment. In this study, think-aloud referred to metacognitive think-aloud, which is “explicit” verbalization of learners' thoughts on the task, such as planning, monitoring, reasoning, and evaluation, because explicit verbalization of thoughts requires more explicit metacognitive processes and thus serves more effectively as a metacognitive strategy in meaning-inference.

Method

Participants

The participants were 39 Korean L1 ESL learners who were enrolled in reading/writing courses in an intensive summer English program in a mid-size university in the U.S. They were randomly assigned to one of the groups, the think-aloud group ($n = 19$), consisting of five males and 14 females, and the control group ($n = 20$), consisting of 11 males and nine females. The mean age was 22.74 ($SD = 2.16$) for the think-aloud group and 22.95 ($SD = 1.73$) for the control group. The mean reading placement test scores are from the *ACT Compass Test*. This test, administered online by *ACT* (originally named American College Testing), is a standardized test, which includes English as a foreign/second language (similar to TOEFL) as well as academic subjects for native-speaking students (similar to the SAT in the United States). The mean reading comprehension scores (maximum 100 points) were 83.16 ($SD = 8.18$) for the think-aloud group and 81.45 ($SD = 7.06$) for the control group. This difference was found to be nonsignificant, $t(37) = -.699, p = .402$. The majority of the students were in an English-speaking country for the first time and had 5 weeks of experience living in the U.S. at the beginning of data collection, except for one student from the think-aloud group who indicated 4 months and one student from the control group who indicated 14 months. All of the participants, whose L1 was Korean, were from the same university in South Korea and were majoring in various fields. In each participant group, more than 50% majored in either engineering or business. The authors were not the participants' teachers. Participation was voluntary.

Materials

Passages for the pretest and posttest were first selected. The criteria for selecting the passages were as follows: (a) they were approximately the same length; (b) the grammar and vocabulary items used in the passages were known to the participants; and (c) the passage topics did not require highly specialized background knowledge. The pretest passage, "When a Young Bird Leaves the Nest" (236 words), was selected from Chern (1993) (see Appendix A) and the posttest passage, "Folk Objects" (253 words), was selected from Hamada and Koda (2011) (see Appendix B). Each passage contained 10 pseudowords, whose meanings the participants were asked to infer. The pseudowords served as unfamiliar L2 words to be learned by the participants. Although the use of pseudowords inherently creates

an unnatural setting, we decided the use of pseudowords was the optimal solution to the design of this study. The use of pseudowords in vocabulary research is a common practice, particularly in experimental studies, because it can minimize the effects of participants' previous vocabulary knowledge and the frequency of real words (Kirsner, 1974). The proportion of known to unknown word coverage for the pretest and posttest passages was 95.76% and 96.05%, respectively. These percentages are above the minimum 95% requirement for meaning-inference to occur, suggested by Liu and Nation (1985).

In order to ensure that the participants would not have difficulty with the grammar or vocabulary items used in the passages except for the 10 pseudowords, their instructors checked the appropriateness. They indicated that the grammar was basic enough and should be known to the participants, but some of the vocabulary items might not be known. Therefore, two weeks prior to the pretest, the participants looked at both the pretest and posttest passages and underlined words whose meanings were unknown to them. This was done as part of their regular class activity and took no longer than 10 minutes. Two Korean L1 graduate students majoring in linguistics/TESOL created a glossary based on the participants' responses, so that the participants would know all of the vocabulary items, except for the pseudowords. Regarding the criteria on the topic, our goal was to not choose a topic that would be intelligible only to those who have specialized knowledge. After discussion with the participants' teachers, we determined that the topics were general enough to be understood by laypeople.

In an effort to examine the comparability of the pretest and posttest passages, ten native speakers of English who were enrolled in degree programs in the same U.S. university read the two passages, inferred the meanings of the pseudowords, and judged the difficulty of the passages. The mean rating for Passage A (pretest) was 2.3 ($SD = 1.16$), and the mean rating for Passage B (posttest) was 2.6 ($SD = .97$), where the scale was 1 very easy, 2 easy, 3 neutral, 4 difficult, and 5 very difficult. The mean rating scores were tested using a two-tailed t test, and the difference was found to be nonsignificant, $t(18) = -.629, p = .538$.

Tasks and Procedures

Pretest

The pretest was administered by the instructors to all of the students in the classroom. The participants read the passage and wrote down the

inferred meanings in both English and Korean. The reason for having them answer both in L2 and L1 was to minimize the influence of L2 proficiency on identifying the correct definition for each pseudoword. The instructors gave a maximum of 30 minutes to complete the pretest, although they noted that most of the participants completed it in 15-20 minutes.

Treatment

The treatment was administered either one day or two days after the pretest. Due to the absence of some of the participants on a field trip, we were unable to schedule the treatment on the same day for all of the participants. The instructors first explained the inference strategies from Nassaji (2003) in class as part of a class session to all of the participants. After the strategy explanation, the control group was asked to stay in the classroom and work on practice inference using the practice passage. They were told to use the strategies introduced, but worked silently. When they completed the practice passage (i.e., wrote down the inferred meanings), they turned it in and were dismissed. As for the think-aloud group, after the strategy explanation provided in class, they worked on practice inference in a separate room, where they received think-aloud instruction from one of the three teaching assistants, Korean L1 graduate students in TESOL who were trained in instructing the think-aloud technique. The participants were asked to use think-aloud as they worked on inference using the practice passage.

The teaching assistants gave the following metacognitive think-aloud instructions in Korean: (a) Say everything that comes across your mind while guessing, but focus on voicing in detail the thought process you used to reach your answer; and (b) Use either Korean or English, whichever is more comfortable for you. The teaching assistants reminded them to keep verbalizing when they stopped doing so for more than 10 seconds. The teaching assistants were not allowed to respond to the participants' questions or utterances during think-aloud. The assistants were told that their role was to ensure that the participants use think-aloud, rather than to help them to infer the meaning of unknown words correctly or to serve as an active listener. The assistants also modeled think-aloud, which included both syntactically complete and incomplete sentences. For both groups, we set the time-on-task as 15 minutes each for the strategy explanation and for the practice inference.

Posttest

The posttest was administered one day after the treatment. Each participant in the think-aloud group completed it individually in a quiet room in the presence of one of the three teaching assistants. Each participant was tested one at a time, and a Sony digital voice recorder was placed on the desk near the participant. After the assistant determined that the participant was relaxed and comfortable using think-aloud, the participant worked on the posttest. The think-aloud group's verbal reports were audio-recorded. The control group worked silently in the classroom on the posttest, administered by their instructors. For the control group, the instructors made sure to create an individual testing atmosphere (not a class activity), and for the think-aloud group, the teaching assistant made sure to be noninterruptive (as if he or she was not present). We were unable to schedule individual rooms for the control group, but we made sure that any difference that might arise would be minimal. As in the pretest, both groups were allowed a maximum of 30 minutes for the posttest.

Analysis Procedures and Results

The following scoring system was used in order to best assess meaning-inference accuracy. Accuracy of the inferred meanings (i.e., participants' definitions of the pseudowords) was judged based on two sources: the original words that were replaced in the passages and the definitions given by the ten native speakers who performed the same task. The participants' definitions were scored following the criteria used in Haynes and Carr (1990). One point was awarded (a) when the definition for a particular pseudoword matched that of the original word; (b) when the definition matched one of the native speakers' definitions; or (c) when the definition was semantically identical (or synonymous) to the original word or one of the native speakers' definitions. A half point was awarded when the inferred meaning was semantically close to the original word or one of the native speakers' definitions. The keys for both passages are presented in Appendix C and Appendix D. Scoring was done by two independent raters who were fluent in both English and Korean (interrater reliability .95), and items whose scoring was disagreed upon were resolved by discussion.

Table 1. Means and Standard Deviations of Pretest and Posttest Scores

Group	Pretest	Posttest
Think-aloud group	3.58 ($SD = 1.72$)	4.66 ($SD = 2.24$)
Control group	3.98 ($SD = 1.31$)	3.83 ($SD = 1.38$)

Table 1 is a summary of the mean scores for the pretest and posttest. The think-aloud group's mean scores were 3.58 ($SD = 1.72$) for the pretest and 4.66 ($SD = 2.24$) for the posttest. The control group's mean scores were 3.98 ($SD = 1.31$) for the pretest and 3.83 ($SD = 1.38$) for the posttest. The Mann-Whitney test was selected for the analysis.

Table 2. Means and Sums of Ranks

Group	Mean/Sum	Pretest	Posttest	Difference
Think-aloud group	Mean	18.84	22.84	23.68
	Sum	358.00	434.00	450.00
Control group	Mean	21.10	17.30	16.50
	Sum	422.00	346.00	330.00

Table 2 is a summary of the means and sums of ranks. The mean and sum of ranks were calculated for the pretest, posttest, and the difference between the two. For the think-aloud group, the mean ranks were 18.84 for the pretest, 22.84 for the posttest, and 23.68 for the difference. For the control group, the mean ranks were 21.10 for the pretest, 17.30 for the posttest, and 16.50 for the difference. The Mann-Whitney U indicated the groups did not differ significantly for the pretest ranks and the posttest ranks, $p = .534$ and $p = .126$, respectively. However, the difference between the pretest and posttest differed significantly between the groups, $U(37) = 120$, $Z = -1.982$, $p = .047$.

To further examine the data, we conducted a qualitative analysis of the think-aloud protocol from the think-aloud group. The verbal reports of the participants who had the top five reading placement test scores ($M = 91$, $SD = 2.35$) and the lowest five reading placement test scores ($M = 75$, $SD = 8.34$) were translated and transcribed in English, then coded into the two types, nonmetacognitive and metacognitive think-aloud. When the verbal report for each pseudoword item included specific information about planning,

monitoring, reasoning, and evaluating, the utterance was coded as metacognitive. When the utterance did not include such information, it was coded as nonmetacognitive. Two raters independently coded the verbal reports (interrater reliability .96), and items whose coding was disagreed upon were resolved by discussion. The protocol analysis excluded one participant's data due to poor audio recording quality.

Table 3. Comparison of the Participants with Higher and Lower Reading Scores

Higher/ Lower	Pretest score	Posttest score	Metacognitive think-aloud	Nonmeta- cognitive think-aloud
Higher	3.4 (<i>SD</i> = 1.52)	5.7 (<i>SD</i> = 3.29)	5.2 (<i>SD</i> = 3.35)	4.8 (<i>SD</i> = 3.35)
Lower	4.2 (<i>SD</i> = 1.35)	4.4 (<i>SD</i> = 2.61)	2.6 (<i>SD</i> = 3.21)	7.4 (<i>SD</i> = 3.21)

Table 3 is a summary of the mean pretest and posttest scores (max. 10) and number of metacognitive think-aloud protocols (max. 10) categorized for the participants with higher (top five) reading placement test scores and lower (lowest five) reading placement test scores. The high scoring group had a larger increase in meaning-inference accuracy in the posttest scores (2.3 points) than the low scoring group did (0.2 points). The high scoring group's protocol contained both metacognitive and nonmetacognitive information in similar amounts (the difference was only 0.4 points), while the low scoring group's protocol had much more nonmetacognitive information (4.8 points more) than metacognitive information.

Discussion

The study investigated whether the use of think-aloud while inferring the meaning of unknown words encountered during reading improves inference accuracy. The results demonstrated that the think-aloud group showed a larger increase in meaning-inference accuracy between the pretest and the posttest than the control group did. As predicted, the think-aloud technique seems to have enhanced the participants' metacognitive processes, which in turn facilitated processes involved in meaning-inference. This finding is in line with the consensus in L1 research (e.g., Loxterman et al., 1994), while inconsistent with the result from the L2 reading comprehension study

(McKeown & Gentilucci, 2007). Although the focus of this study was not to test the reactivity issue, the current results suggest that think-aloud was positively reactive in meaning-inference during reading. Similar results, an increase in accuracy, were reported in L2 grammar learning studies (e.g., Rossomondo, 2007; Sanz et al., 2009), yet most L2 reading comprehension studies showed either no reactivity (e.g., Leow & Morgan-Short, 2004) or negative reactivity (e.g., Bowles & Leow, 2005). These findings offer a possible interpretation that meaning-inference involves cognitive processes, perhaps processes involved in problem-solving, more similar to grammar learning than to reading comprehension. Needless to say, further research is warranted to confirm the current findings and interpretations, given the limited amount of research in think-aloud technique and L2 learning, including reading comprehension and meaning-inference. In particular, more findings are necessary to verify whether the role of think-aloud in L1 learning tasks differs from that in L2 learning tasks and whether think-aloud plays a different role in the different learning activities, reading comprehension and meaning-inference.

The follow-up qualitative analysis on the protocol from the think-aloud group also offers interesting insights. The group with higher reading scores seemed to show a larger positive effect from the use of think-aloud in meaning-inference than the group with lower reading scores. Of particular interest is that the low-scoring group's protocol contained much more non-metacognitive information than the high scoring group's protocol did. These results appear to suggest that the use of think-aloud relates to meaning-inference accuracy and the relationship also varies according to individual learners' English proficiency. Although an effort was made to ensure that the two groups, the think-aloud and control groups, were comparable in terms of age and English proficiency, a more strict homogeneity of the participants will be necessary in future studies.

For those who used nonmetacognitive think-aloud more frequently, the overwhelmingly most common characteristic of the verbal protocol was to simply read and reread a portion of the passage. Following is one such example. Because the participant read from the passage, the entire utterance was in English.

“Some individuals ricate with objects as though they were people. . . . Some individuals ricate with objects as though they were people.”

In contrast, those who used metacognitive think-aloud revealed reasoning based mainly on the three pieces of information, morpho-syntactic structure, world knowledge, and discourse knowledge, which Nassaji (2003) reported lead to meaning-inference success. In the examples below, the text portions indicated by single quotation marks were uttered in English, and the rest was uttered in Korean.

Use of morpho-syntactic structure

“It is a noun if seeing next sentence. . . . ‘They may serve as symbols for social class, kede, or ethnicity’. . . ‘social class . . . object display’ . . . because this is a plural, should be many. . .”

“‘They may serve as symbols for social class, kede, or ethnicity’. . . ‘kede’. . . it is parallel . . . so . . . could be the same meaning as ‘social class,’ and ‘ethnicity’ . . . then religion?”

Use of world knowledge

“It has been taded that in Utah, one can find driveways lined with wheels, and gates built from commercial objects’ . . . um . . . this is something we can see in Utah . . .”

Use of discourse knowledge

“Although mailboxes must follow official standards of measurement’ . . . so . . . they should follow the standards, but they do not . . . so it means change.”

Because metacognitive think-aloud requires a deeper level of processing (Ericsson & Simon, 1993), it makes sense that those who were unable to perform metacognitive think-aloud were only able to read aloud a portion of the passage, only one of the tasks involved in meaning-inference while reading. This observation, although qualitative in nature, seems to indicate the interconnection between English proficiency, including proficiency in reading, the quality of think-aloud, and metacognitive strategies.

Conclusion

This study explored the effectiveness of think-aloud in L2 meaning-inference during reading with college-level ESL students. The findings from the study suggest that think-aloud helps facilitate accurate meaning-inference, as shown by the increase in inference accuracy in the student group who

used the think-aloud technique. A follow-up analysis on the protocol from the think-aloud group seems to imply that the individual participants' English proficiency relates to the use of metacognitive strategies, the type of information uttered during think-aloud, and meaning-inference success. Given that this observation is based on a qualitative analysis, further research is warranted to provide any definitive conclusions, addressing the limitations of participant grouping (to have a more homogeneous group in terms of English proficiency) and task procedure (to have the control group given the posttest individually in a separate room, as was the think-aloud group).

Finally, pedagogical implications are addressed. It is recommended that meaning-inference instruction incorporate think-aloud strategy instruction as well. However, teachers and students need to be aware of false causality—simply using think-aloud does not necessarily lead to greater success in meaning-inference. The quality of think-aloud needs to be focused on during instruction in order to ensure that students' think-aloud actually reflects their metacognition and serves as a metacognitive strategy. Metacognitive strategies typically include planning, reasoning, monitoring, and evaluation of one's learning activity (Purpura, 1997). Without the existence of these characteristic components in learners' verbal reports, the effectiveness of think-aloud will be reduced. While the researchers explored the instructional implementation of think-aloud, they advise teachers and students to incorporate think-aloud in L2 meaning-inference with some caution. It is also recommended that teachers incorporate meaning-inference instruction only after knowing whether their students' English proficiency level is high enough to be conducive to correct inferences.

Acknowledgements

We thank the students who participated in this study. Thanks also to Mary Theresa Seig, Kathy Ramos, and Kathleen Ulrey for their assistance in recruiting participants and Chaeyoung Kim Lee and Boyoung Roh for their assistance in data collection and analysis. Permission has been obtained to reproduce the contents in Appendix A from Ablex Publishing and Appendix B from Elsevier.

Megumi Hamada teaches at Ball State University. Her specialty is second language acquisition. *Chaehee Park* teaches in South Korea. His specialty is second language spelling and reading.

References

- Baumann, J. F., Seifert-Kessell, N., & Jones, L. A. (1992). Effect of think-aloud instruction on elementary students' comprehension monitoring abilities. *Journal of Reading Behavior, 24*, 143-172.
- Bensoussan, M., & Laufer, B. (1984). Lexical guessing in context in EFL reading comprehension. *Journal of Research in Reading, 7*, 15-32.
- Bereiter, C., & Bird, M. (1985). Use of thinking aloud in identification and teaching of reading comprehension strategies. *Cognition and Instruction, 2*, 131-156.
- Berne, J. (2004). Think-aloud protocol and adult learners. *Adult Basic Education, 14*, 153-173.
- Block, E. L. (1992). See how they read: Comprehension monitoring of L1 and L2 readers. *TESOL Quarterly, 26*, 319-343.
- Bowles, M. A. (2008). Task type and reactivity of verbal reports in SLA: A first look at a L2 task other than reading. *Studies in Second Language Acquisition, 30*, 359-387.
- Bowles, M. A. (2010). *The think-aloud controversy in second language research*. New York: Routledge.
- Bowles, M. A., & Leow, R. P. (2005). Reactivity and type of verbal report in SLA research methodology: Expanding the scope of investigation. *Studies in Second Language Acquisition, 27*, 415-440.
- Carrell, P. L. (1989). Metacognitive awareness and second language reading. *The Modern Language Journal, 73*, 121-134.
- Casanave, C. P. (1988). Comprehension monitoring in ESL reading: A neglected essential. *TESOL Quarterly, 22*, 283-302.
- Chern, C.-L. (1993). Chinese students' word-solving strategies in reading in English. In T. Huckin, M. Haynes, & J. Coady (Eds.), *Second language reading and vocabulary learning* (pp. 67-85). Norwood, NJ: Ablex.
- Egi, T. (2008). Investigating stimulated recall as a cognitive measure: Reactivity and verbal reports in SLA research methodology. *Language Awareness, 17*, 212-228.
- Ellis, R. (2001). Investigating form-focused instruction. *Language Learning, 51*, 1-46.
- Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data* (Rev. ed.). Cambridge, MA: MIT Press.
- Flavell, J. H. (1976). Metacognitive aspects of problem solving. In L. B. Resnick (Ed.), *The nature of intelligence* (pp. 231-235). Hillsdale, NJ: Erlbaum.
- Goo, J. (2010). Working memory and reactivity. *Language Learning, 60*, 712-752.

- Haastrup, K. (1991). *Lexical inferencing procedures or talking about words: Receptive procedures in foreign language learning with special reference to English*. Tübingen, Germany: Gunter Narr.
- Hamada, M., & Koda, K. (2011). Similarity and difference in learning L2 word-form. *System*, 39, 500-509.
- Haynes, M. (1993). Patterns and perils of guessing in second language reading. In T. Huckin, M. Haynes, & J. Coady (Eds.), *Second language reading and vocabulary learning* (pp. 46-64). Norwood, NJ: Ablex.
- Haynes, M., & Carr, T. H. (1990). Writing system background and second language reading: A component skills analysis of English reading by native speaker-readers of Chinese. In T. H. Carr & B. A. Levy (Eds.), *Reading and its development: Component skills approaches* (pp. 375-421). San Diego, CA: Academic Press.
- Huckin, T., & Bloch, J. (1993). Strategies for inferring word meaning in context. In T. Huckin, M. Haynes, & J. Coady (Eds.), *Second language reading and vocabulary learning* (pp. 153-178). Norwood, NJ: Ablex.
- Huckin, T., & Coady, J. (1999). Incidental vocabulary acquisition in a second language: A review. *Studies in Second Language Acquisition*, 21, 181-193.
- Jourdenais, R. (2001). Cognition, instruction and protocol analysis. In P. Robinson (Ed.), *Attention and awareness in foreign language learning* (pp. 183-216). Honolulu: University of Hawai'i, Second Language Teaching and Curriculum Center.
- Kaivanpanah, S., & Alavi, S. M. (2008). The role of linguistic knowledge in word-meaning inferencing. *System*, 36, 172-195.
- Kelly, P. (1990). Guessing: No substitute for systematic learning of lexis. *System*, 18, 199-207.
- Kirsner, K. (1974). Modality differences in recognition memory for words and their attributes. *The Journal of Experimental Psychology*, 102, 579-584.
- Koda, K. (2005). *Insights into second language reading: A cross-linguistic approach*. Cambridge: Cambridge University Press.
- Kucan, L., & Beck, I. L. (1997). Thinking aloud and reading comprehension research: Inquiry, instruction, and social interaction. *Review of Educational Research*, 67, 271-299.
- Laing, S. P., & Kamhi, A. G. (2002). The use of think-aloud protocols to compare inferencing abilities in average and below-average readers. *Journal of Learning Disabilities*, 35, 436-447.
- Leow, R. P., & Morgan-Short, K. (2004). To think aloud or not to think aloud: The issue of reactivity in SLA research methodology. *Studies in Second Language Acquisition*, 26, 35-57.

- Liu, N., & Nation, I. S. P. (1985). Factors affecting guessing vocabulary in context, *RELC Journal*, 16, 33-42.
- Loxterman, J. A., Beck, I. L., & McKeown, M. G. (1994). The effects of thinking aloud during reading on students' comprehension of more or less coherent text. *Reading Research Quarterly*, 29, 353-367.
- Maria, K., & Hathaway, K. (1993). Using think alouds with teachers to develop awareness of reading strategies. *Journal of Reading*, 37, 12-18.
- McKeown, R. G., & Gentilucci, J. L. (2007). Think-aloud strategy: Metacognitive development and monitoring comprehension in the middle school second-language classroom. *Journal of Adolescent & Adult Literacy*, 51, 136-147.
- Nassaji, H. (2003). L2 vocabulary learning from context: Strategies, knowledge sources, and their relationship with success in L2 lexical inferencing. *TESOL Quarterly*, 37, 645-670.
- Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.
- Oster, L. (2001). Using the think-aloud for reading instruction. *The Reading Teacher*, 55, 64-69.
- Paribakht, S., & Wesche, M. (1999). Reading and "incidental" L2 vocabulary acquisition: An introspective study of lexical inferencing. *Studies in Second Language Acquisition*, 21, 195-224.
- Pressley, M., & Afflerbach, P. (1995). *Verbal protocols of reading: The nature of the constructively responsive reading*. Hillsdale, NJ: Erlbaum.
- Purpura, J. (1997). An analysis of the relationships between test takers' cognitive and metacognitive strategy use and second language test performance. *Language Learning*, 47, 289-325.
- Rossomondo, A. E. (2007). The role of lexical temporal indicators and text interaction format in the incidental acquisition of the Spanish future tense. *Studies in Second Language Acquisition*, 29, 39-66.
- Sachs, R., & Polio, C. (2007). Learners' uses of two types of written feedback on a L2 writing revision task. *Studies in Second Language Acquisition*, 29, 67-100.
- Sachs, R., & Suh, B. R. (2007). Textually enhanced recasts, leaner awareness, and L2 outcomes in synchronous computer-mediated interaction. In A. Mackey (Ed.), *Conversational interaction in second language acquisition: A collection of empirical studies* (pp. 197-228). Oxford: Oxford University Press.
- Sanz, C., Lin, H.-J., Lado, B., Bowden, H. W., & Stafford, C. A. (2009). Concurrent verbalizations, pedagogical conditions, and reactivity: Two call studies. *Language Learning*, 59, 33-71.

- Sternberg, R. J. (1987). Most vocabulary is learned from context. In M. G. McKeown & M. E. Curtis (Eds.), *The nature of vocabulary acquisition* (pp. 89-105). Hillsdale, NJ: Erlbaum.
- Walker, B. J. (2005). Thinking aloud: Struggling readers often require more than a model. *The Reading Teacher*, 58, 688-692.
- Yoshida, M. (2008). Think-aloud protocols and type of reading task: The issue of reactivity in L2 reading research. In M. Bowles, R. Foote, S. Perpiñán, & R. Bhatt (Eds.), *Selected proceedings of the 2007 Second Language Research Forum* (pp. 109-209). Somerville, MA: Cascadilla Proceedings Project.

Appendix A

Passage A—When a Young Bird Leaves the Nest

Like people, young birds go through a difficult transition when it's time to strike out on their own. The fledgling must be (1) glurked while learning to feed itself. It must be protected while learning to fly. In some species, fledglings must even be (2) moxed by their parents during their first autumn migration.

In most cases, a young bird (3) tidly returns once it leaves the nest. But there are some (4) padons. The youth of certain kinds of woodpeckers, wrens and swallows fly back to the nest to sleep. Similarly, some eagles and large hawks (5) firk home for weeks to feed until they learn how to catch their own (6) pum.

When it comes to (7) snerdling, however, few fledglings need any lessons. Fifty years ago, a German scientist names J. Grohmann raised some young pigeons in narrow tubes that prevented them from moving their (8) lurds. At the same time he allowed another group of pigeons of the same age to be raised by their (9) medlons in a nest in the normal way, exercising their wings vigorously. When the two groups of pigeons were mature enough, Grohamm took them out into the open and tossed them into the air. Surprisingly, the pigeons raised in the tubes flew away as strongly as the ones that had been unrestrained in the nest. Grohamm thus proved that the instinctive (10) grumility to fly develops in young birds with or without the opportunity to practice.

Appendix B

Passage B—Folk Objects

The relationships that objects have with their human creators and owners are recognizable. Object forms show human characteristics, for example, chairs are (1) noked as having legs, lamps as having necks, and clocks as having faces. Some individuals (2) ricate with objects as though they were people. They give them names, talk to them, and decorate or “dress” them. In American culture, for example, cars are regularly named or personalized with special license plates or paintings. They may be praised for good performance or cursed for bad. Some (3) beek consider the purchase of new mats, covers, or ornaments as buying “gifts” for their cars. So, humans express their own ideas and (4) hoakings through objects and see them as reflections of themselves.

Objects can be used for display (5) clibes of their human creators and owners. They may serve as symbols for social class, (6) kede, or ethnicity. A (7) deany example of object display can be found in front of houses. It has been (8) taded that in Utah, one can find driveways lined with wheels, and gates built from commercial objects. Although mailboxes must follow official standards of measurement, owners (9) soanalyze them. The mailboxes are converted into symbols of personal, occupational, or regional identity. Cowboys and horses (10) ficed from steel are put on the tops of mailboxes. The bottom is built from milkcans and wheels. Many mailboxes have iron chains built into supports, and bent to form initials or abstract shapes. By using objects, humans display their characteristics within what they believe to be a more uniform culture.

Appendix C

Passage A Answer Key

Pseudoword	Original word (1 point)	Native speaker answer other than the original word (1 point)	Synonym of the original word or a native speaker answer (1 point)	Meaning is partially correct (0.5 point)
(1) glurked	fed	aided/assisted, guided, given food, protected, helped	prevented 보살핌을 받다	raised (기르다)

(2) moxed	accompanied	forced, led, carried, helped along, supervised, watched over, guided, helped, coaxed, urged, assisted	instruct (지도하다, 가르치다) to follow (따라가다), to manage	depend on (의지하다)
(3) tidly	rarely	seldom, never, quickly, hardly ever, almost never		
(4) padons	exceptions	problems, issues, cases, situations, birds that return to the nest even though they aren't expected to	except pattern tendency (경향)	types of behavior (행동양식) habit (습성), rule (규칙)
(5) firik	return	come, stay, visit, leave		
(6) pum	prey	food	meal	
(7) snerdling	flying	instincts		independence (독립) moving (이동하다)
(8) lurds	wings	bodies		
(9) medlons	parents	family, mother(s)		
(10) grumlity	ability	instinct, urge, desire, need	skills nature nature (성질)	power, will (힘, 의지) behavior, action (행동) sensitivity (감각) characteristic (특성, 특징) way to survive (생존방법)

Appendix D

Passage B Answer Key

Pseudoword	Original word (1 point)	Native speaker answer other than the original word (1 point)	Synonym of the original word or a native speaker answer (1 point)	Meaning is partially correct (0.5 point)
(1) noked	described	said to have, deemed, thought as, known, defined, recognized, tables, talked about	symbolize recognize represent (나타내어진다)	similar, seems (같다, 비슷하다) call as~ (~라고 부르다) seen (~처럼 보인다)
(2) ricate	interact	spend time, talk, identify, bond, relate		consider (간주하다), think, recognize express (나타내다)
(3) beek	individuals	owners, enthusiasts, people, supporters, drivers	car lover (차애호가) 사람	consumer, buyer (구매자, 소비자) mania (매니아)
(4) hoakings	feelings	identifications, tastes/styles, opinions, personalities, desires, likes	신념, 성격, 모토, 기호 thinking (생각), 가치관	image (이미지)
(5) clibes	purposes	characteristics, places, a part of, a piece of a whole, facets, pieces, boards, symbols, trophies	조각, 부분	thinking, emotion, opinion, idea 생각, 의견, 문화

(6) kede	occupation	age, religion, race, gender, membership, identity, culture	개인적 특성, personal characteristic, identity	tradition, rich (부), 명성(fame), status (지위), economic status, honor (명예), social status (계급), authority (권위), money (돈)
(7) deany	contemporary	good, common, great, perfect, prime	looking easily (쉽게 찾을 수 있는)	representative (대표적인), most (대부분의) 분명한, 적절한, 확실한, 알맞은, 전형적인, similar (비슷한), specific (구체적인)
(8) taded	reported	noticed, seen, said, stated, found, recorded, shown	showed, indicate, find, 나타내다	발생한, exist, 적용되다
(9) soanalize	personalize	decorate, dress them up, specialize	hope to make it specific (~특별하게 하고 싶다) identify, distinguish (개성화하다)	change
(10) ficed	made	cut, created, shaped	built, carve, form	꾸미다, 문양되다