139
196
THS

THESIS
I
2006

This is to certify that the
thesis entitled

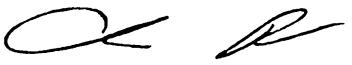
ANOTHER LOOK AT REACTIVITY IN L2 THINK-ALOUD PROTOCOLS:A REPLICATION STUDY

presented by

JIAWEN WANG

has been accepted towards fulfillment
of the requirements for the

M. A. degree in TESOL



Major Professor's Signature

August 25, 2005

Date

LIBRARY
Michigan State
University

MSU is an Affirmative Action/Equal Opportunity Institution

PLACE IN RETURN BOX to remove this checkout from your record.
TO AVOID FINES return on or before date due.
MAY BE RECALLED with earlier due date if requested.

DATE DUE	DATE DUE	DATE DUE

2/05 c:/CIRC/DateDue.indd-p.15

ANOTHER LOOK AT REACTIVITY IN L2 THINK-ALOUD PROTOCOLS: A REPLICATION STUDY

By

Jiawen Wang

A THESIS

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

MASTER OF ARTS

Department of Linguistics and Germanic, Slavic, Asian, and African Languages

2005

ABSTRACT

ANOTHER LOOK AT REACTIVITY IN L2 THINK-ALOUD PROTOCOLS:
A REPLICATION STUDY

By

Jiawen Wang

This current study replicated Leow and Morgan-Short (2004) with thirty (fifteen in the experimental/think-aloud group and fifteen in the control/nonthink-aloud group) native Chinese students with higher proficiency in a second language (English), using a reading passage and the same three types of assessment tasks, which included a multiple-choice comprehension task, a recognition task to determine the learners' intake of phrasal verbs, and a controlled written production task. The nonthink-aloud group in this study outperformed the think-aloud group in the reading comprehension task and the target language recognition task, indicating that thinking aloud while performing a reading task seemed to have detrimental effects on learners' comprehension and intake, but did not seem to affect controlled written production. This study also qualitatively examined the think-aloud protocols as a possible influence on the presence or absence of reactivity. The conclusion is that, the think-aloud protocol is not simply reactive or nonreactive. It is the result of dynamic interactions between several factors with L2 learners' translation strategy as only one of them. It is suggested that more systematic research (including replication) is necessary to have a clearer and more comprehensive picture of the whole reactivity issue regarding think-aloud protocols.

To

My wife and my child

Who have been with me toward my goal

ACKNOWLEDGEMENTS

The completion of this study would be impossible without the responsible and professional instructions from Dr. Charlene Polio and Dr. Susan Gass, who awakened my interest in research and made me feel I made a right choice to come to Michigan State University for my degree in MA TESOL. I also owe special thanks to Robin Revette Roots, Nigel Caplan, and Matthew Rynbrandt, not only for their help in editing this thesis, which was particularly important for me as a writer with English as a second language, but also for their brief comments and discussions in the process. Finally, but not last, I should thank my participants for their contribution of time and insights to this study. Despite great help from the above mentioned professors, colleagues, and participants, whatever mistakes or errors are of my own responsibility.

TABLE OF CONTENTS

LIST OF TABLES

Another Look at Reactivity in L2 Think-Aloud Protocols: A Replication Study

The Think-aloud Protocol

Verbal reports and protocol analysis represent one evolution of the human habit of asking people to share their thoughts into a useful form of scientific inquiry, and the last two decades have witnessed the burgeoning use of protocol analysis to investigate acts of cognition, response, and reading related phenomena (Afflerbach, 2002). As Ericsson and Simon (1987, p. 32) defined, "To obtain verbal reports, as new information (thoughts) enters attention, the participants should verbalize the corresponding thought or thoughts...the new incoming information is maintained in attention until the corresponding verbalization of it is completed."

More general than the concept of verbal reports, introspective reports are generally considered to differ along a number of dimensions: currency (i.e., time frame), form (i.e., oral, written), task type (i.e., think-aloud, talk-aloud, retrospective), and support provided to the participants in reporting (Gass and Mackey, 2000, pp. 13-14). Cohen (2000) has subcategorized verbal reports into three types based on the nature of the content: (a) self-report, (b) self-observation, either introspectively or retrospectively and not so general as self-report, and (c) self-revelation, "think-aloud", stream-of-consciousness disclosure of thought processes while information is being attended to. Ericsson and Simon (1993)

categorized verbal reports as either concurrent or retrospective based on the temporal frame in which the reports are collected. Introspective reports are made while a participant is performing a task. Retrospective reports are made in a short time, usually immediately, after a task or part of a task has been performed. In addition, Ericsson and Simon also made a major distinction between the instruction to verbalize thoughts per se and instructions to verbalize specific information, such as reasons and explanations about the participants' thinking process, with the former similar to Cohen's (2000) self-revelation and with the latter similar to Cohen's self-observation. For the purposes of their studies in the SLA area, Leow and Morgan-Short (2004) and Bowles and Leow (2005) referred to verbalizations per se as nonmetalinguistic and those requiring additional specific information as metalinguistic. For example, a typical instruction for a nonmetalinguistic procotol would be to ask participants to think their thoughts aloud while reading an article and answering the questions, that is, to say whatever passes through their mind during the process of completing their task. A typical instruction for a metalinguistic protocol would be to ask participants to "verbalize every thought and every detail of your thought process, including what information you are looking at, what thoughts you are having about any piece of information, how you evaluate different pieces of information, *and why*" (Bowles and Leow, 2005, p. 426).

Having also identified another term for verbal reporting, *process tracing* in Shavelson, Webb, and Burstein (1986), which lists three types of verbal reporting (think-aloud or talk-aloud during a task, thinking about a previously performed task, and prompted interviews), Gass and Mackey (2000) suggested, "Despite different terminology, verbal reporting can be seen as gathering data

by asking individuals to vocalize what is going through their minds as they are solving a problem or performing a task" (p.13).

The think-aloud protocol is generally considered to be introspective (concurrent) and non-metalinguistic because "the standard method for getting participants to verbalize their thoughts concurrently is to instruct them to 'think-aloud'" (Ericsson and Simon, 1993, p. xiii) and "the 'think-aloud' instruction explicitly warns the participants against explanation and verbal description" (p. xiv). Although the term *retrospective think-aloud* may sometimes appear, as in Anderson (1989) and Fraser (1999), it does not mean what is normally understood for think-aloud protocols. For example, although Fraser (1999) reported asking the participants to do their retrospective think-aloud, what the participants did was to engage in an oral interview after the task, responding to probes like "What did you do and think about when you first saw [the word] 'X'?" (p. 228). This is the retrospective protocol, not the idea of think-aloud in the commonly understood sense of the term.

SLA Research That Applied Think-aloud Protocol

Although the use of verbal reports to investigate cognitive processes in various areas of psychology, cognitive science, and education has a longer history, their use in SLA research has also had a history of several decades. The topics researched include vocabulary, reading, writing, L2 test-taking, strategy use, grammaticality judging, and translation amongst other areas. Gass and Mackey (2000, p.29) provided a table listing a

sampling of second language studies using introspection. Table 1 has been adapted from Gass and Mackey (2000) to show only those studies that used think-aloud protocols. Note that the think-aloud mentioned here may have various foci including metalinguistic and non-metalinguistic, and introspective and retrospective and these studies were listed here only because Gass and Mackey had listed them as using think-aloud protocols no matter what adaptations the related researchers had made to this technique. Despite this, in this table, the studies related to reading tasks are in bold and more details about their think-aloud protocols are reviewed and briefed below (in a separate Table 2) because they are more closely related to the tasks used to elicit data in this current study and Leow and Morgan-Short (2004).

Block (1986) used think-aloud protocols to examine the comprehension strategies used by 9 college-level students---both native speakers of English (3) and nonnative speakers (6)---enrolled in remedial reading classes as they read material from a college textbook. "Poor readers" (p. 463) were used because they were believed not to have attained the degree of automaticity found in fluent readers, to be more aware of how they solved the problems they encountered as they read, and therefore suitable for the use of think-aloud protocols. The ESL participants were judged by their reading teachers to be fairly fluent in English, so the reporting language was English (L2). Two passages from a college textbook were used as reading materials. The participants were asked "to report exactly what they were thinking while reading and were cautioned against trying to explain or analyze their thoughts" (p. 469). Apparently this is concurrent nonmetalinguistic think-

Table 1. Second Language Studies Using Think-aloud[1]

Author	Year	Type of Data
Abraham and Vann	1996	L2 test-taking
Anderson	1989	L2 test taking
Alanen, R.	**1995**	**Reading**
Block	**1986**	**Reading**
Brice	1995	writing
Cavalcanti	**1987**	**Reading**
Chern	1993	Vocabulary.
Cohen	1994	L2 test taking
Cohen and Cavalcanti	1990,1987	writing
Cohen, Weaver and Li	1995	strategy use
Davies and Kaplan	1998	Grammaticality judgments
Enkvist	1995	Translation
Fach and Kasper	1986	Translation
Feldman and Stemmer	1987	L2 Test-taking
Gerloff	1987	Translation
Goass, zhang and Lantolf	1994	Grammaticality judgments
Gu	1994	Vocabulary
Haastrup	1987	Vocabulary
Hoscher and Mohle	1987	Translation
Hosenfeld, C.	**1976, 1977, 1979, 1984**	**grammar (1976), reading (1977, 1979, 1984)**
Huckin and Bloch	1993	Vocabulary
Jones and Tetroe	1987	Writing
Jourdenais, Ota, Stauffer, Boyson, and Doughty	1995	Linguistic knowledge
Kern	**1994**	**Reading**
Krings	1987	Translation
Lay	1982	Writing
Neubach and Cohen	1988	Dictionary use
Paribakht and Wesche	1999	Vocabulary
Raimes	1985	Writing
Robinson	1991	Pragmatics/Speech acts
Skibniewski	1990	Writing
Stemmer	1991	L2 test-taking
Swain and Lapkin	1995	Writing
Tomitch	**1999**	**Reading**
Vignola	1995	Writing
Zimmermann and Schneider	1987	Vocabulary

Table 2. Some studies that applied think-aloud protocols to research with reading tasks.

Studies	Field	Participants	Materials	Nonthink-aloud tasks combined with	Reporting Language	Way of think-aloud	About think-aloud
Block (1986)	Comprehension strategies	9 entering freshmen (6 NNS, 3NS of English) enrolled in remedial reading courses	Two passages from a textbook(both in English)	Retelling and multiple-choice test. Between participants.	English for both groups	Read silently but think aloud after each sentence for one passage and after each paragraph for another passage; no trying to explain or analyze.	1. Same amount of time for NS and NNS. 2. Discomfort in thinking aloud after each sentence. 3. Think-aloud may assist awareness of strategy resources.
Cavalcanti (1987)	Pragmatic interpretation problems	NS of English and Brazilian EFL readers. Number and proficiency level was not mentioned.	An English passage and a Portuguese passage.	CLOZE task. Between participants.	Not mentioned.	Read silently but think aloud whenever there is a pause.	1. Training is necessary. 2. Longer pauses. 3. Indirect representation of reading process.

Hosenfeld (1984)	Reading strategies (focused on word meaning)	2 separate case studies with ninth grade students of French and Spanish	Unassigned passages from textbooks	Repeated measure comparing protocols within participants. No extra tasks.	L1.	Concurrent. Used with students who translate	1. Causing more translation? 2. Changing participants' strategies?
Kern (1994)	Mental translation	51 intermediate French students in high, middle, and low reading ability groups	2 French passages from intermediate-level textbook	Repeated measure comparing protocols within participants and between participants. No extra tasks.	Not stipulated in instruction. Yet all participants but one reported in English (L1)	Combined with "sentence by sentence" text interview procedure.	1. Believed to be good for translation process research. 2. Normal reading process may be distorted.
Leow (2001)	Attention	38 first-year college-level participants of Spanish	1 modified article from a Spanish weekly magazine	3 after-reading assessment tasks	Not mentioned.	Concurrent, non-metalinguistic.	1. Believed to be robust in providing online data of reading process. 2. No other specific discussion.

aloud. However, a variation from what is normally known about concurrent think-aloud was that the participants were told to offer think-aloud responses only after silent reading of each sentence (for the first passage) and after silent reading of each paragraph. This kind of think-aloud takes on some characteristics of retrospective verbal report. Block designed a retelling task and a 20-item multiple choice test to measure the amount of information understood and remembered by the participants, and related strategy use revealed in the think-aloud protocol to the measures of memory and comprehension. In addition to the discussion of strategies used by the participants, Block also discussed issues related to the think-aloud protocol. First, the time used by the ESL participants and the native participants was quite similar and this suggested that "all readers were able to perform the think-aloud task" (p. 475) and that ESL readers appeared to have performed the task with as much ease or discomfort as native speakers. Second, both some ESL readers and some native speakers complained more about the requirement to respond after reading each sentence and less about the requirement to respond after reading each paragraph. Third, think-alouds may be an important learning tool because several participants reported how the task of think-aloud seemed to have made the participants aware of what they were doing and understood and therefore aware of the strategic resources they might turn to. Therefore, although the purpose of Block's (1986) study was strategy use in reading, it also reported both detrimental and facilitative effects of an adaptation of think-aloud protocols.

Another adaptation of think-aloud protocol was made by Cavalcanti (1987) to examine areas of pragmatic interpretation problems encountered by FL readers in

tackling the introduction to an academic paper. "*Pragmatic interpretation* refers to the striving for equilibrium between reader-relevance and text salience" (p. 237). Cavalcanti named her verbal protocols "pause protocols" (p. 238) because the participants were asked to read silently first and think aloud whenever they noticed a pause in the reading process, which is a little different from Block (1986), who required the participants to think aloud after each sentence or paragraph. Cavalcanti also explained why the participants were not asked to think aloud while reading: A pilot study using this technique indicated that the participants usually ended up reading large chunks of text and then self-reporting. The participants thought aloud on an English text and then on a Portuguese text that served the purpose of a comparison measure. The pause protocols were also combined with four control measures (title study task for content anticipation, interventionist procedure for pause occurrence, oral summary for comprehension, and selection of key lexical items to check basis inter-participant agreement on key lexical items) taken at various predetermined stages of reading. Therefore, Cavalcanti's adaptation of think-aloud protocol was in a sense quite complex with many factors involved. In addition to some implications about the observational findings with respect to FL reader-text interaction, Cavalcanti discussed pause protocol's advantage as a promising attempt to capture the ongoing reading process and its limitations, such as demanding training that helps readers to be aware of their own processing of information, entailing pauses which may be longer than in a real reading situation, and perhaps more seriously, resulting in an over-elaboration that leads to data only indirectly representing the reading process.

Hosenfeld's line of research (1976, 1977, 1979, 1984) applied think-aloud protocols to identify successful and unsuccessful foreign language learners' reading strategies related to the solution of word-meaning problems and to meaning retention while docoding. Here we focus on Hosenfeld (1984) as an example. Hosenfeld asked the following research question: can unsuccessful readers acquire the strategies of successful readers? Hosenfeld reported two case studies, one with a fourteen-year-old girl in a level two French class, and the other with a fourteen-year-old boy struggling with Spanish. Think-aloud was applied to diagnose each subject's strategies in reading an unassigned passage from their textbooks. Although the two cases were about different aspects of reading strategies (because the two learners had different difficulties), in both cases, the participants made marked improvement in applying more reading strategies including some new ones after the remedial session during which Hosenfeld compared what he found from the participants' think-aloud protocols with the strategies used by successful learners. Therefore, Hosenfeld's work suggested the strength of think-aloud protocols in gaining insight into readers' thinking process. One of the principles suggested by Hosenfeld in using think-aloud protocol was to use it "with students who translate" (p. 232). Hosenfeld's own concern with think-aloud protocols was also in translation: Does the thinking aloud cause some students to translate more than they normally do? Another concern is whether the method changes students' strategies in other ways.

In agreement at a different angle with Hosenfeld's (1984) suggestion of using think-aloud protocol with those learners who translate, Kern (1994) indeed applied think-aloud protocols to the research into the role of mental translation of learners in second language

reading. Kern gave all 51 participant students enrolled in French 3 (third-semester university students, in high, middle, and low reading ability groups) an individual "reading task interview" (p. 443) twice to assess their use of translation and other mental procedures when reading French texts, once at the beginning of the semester and again at the end. Similarly to Block (1986), Kern also asked the participants to think-aloud sentence by sentence; however, two differences are that Kern's participants did not have to wait until the end of a sentence to think aloud and that Kern only presented to the participants one new sentence at a time instead of presenting the whole passage. The participants were free to return to earlier sections of the text for clarification. The investigator's role was to prompt participants by asking "what are you thinking now?" following each sentence. Kern also designed a recall protocol task at the end of the passage and after taking the passage away from the participants, asking them to identify the main idea of the passage for the purpose of associating translation reports with comprehension at a later stage. Making use of the think-aloud protocol, Kern identified the specific contexts in which participants relied on translation and analyzed the functional benefits and strategic uses of translation. While arguing for the appropriateness of think-aloud protocols for research into translation in reading process, Kern also showed concern that the combination of think-aloud with the procedure of sentence-by-sentence presentation might distort the normal reading task.

The above reviewed studies seem to combine some other features with their so-called think-aloud protocols in investigating different aspects of the reading process. The line of research studying attentional aspects of the second language reading, e.g. Alenan

(1995), Leow (1997, 1998a, 1998b, 2001), Rosa and O'Neill (1999), and Rott (1999), is relatively more homogenous in the application of think-aloud in its sense of concurrent and nonmetalinguistic verbal reports. To address the effects of formal instruction or exposure, most SLA studies have employed a pretest, instruction-exposure, post-test research design to draw conclusions about the benefits or lack thereof of such instruction or exposure on learners' subsequent processing of the second or foreign language data. The aforementioned studies began to address the methodological issue of internal validity of the traditional research design by employing verbal reports to gather concurrent data to measure the role of attention while learners interacted with the L2 data. Among the few studies that applied think-aloud protocols to the study of awareness, attention, and intake, Rosa and O'Neil (1999) claimed that they followed Leow's (1997, 1998a, 1998b) method of using think-aloud protocols to research their topics. This method can be explicated by analyzing Leow's later (2001) study to understand how think-aloud protocols were used in previous attentional studies. Leow (2001) asked 38 first-year college-level participants (21 in the experimental/enhanced group and 17 in the control/unenhanced group) to do think-aloud while reading a modified Spanish article with the formal imperative in Spanish as the target linguistic form and completing three subsequent assessment tasks aimed to measure intake, written production of the targeted linguistic forms, and the comprehension of the article. To think aloud, the participants had to put on headphones and, "as naturally as they could...clearly speak aloud their thoughts throughout the entire experiment, that is, while reading the article and completing the tasks" (p. 501). This kind of think-aloud is concurrent and non-metalinguistic. Leow also used the term online in the sense of concurrent. After defining and tabulating noticing in the protocol, Leow

compared the results with those of the assessment tasks and discussed the relations between enhanced written input, reported noticing, intake, and written production. In this way, think-aloud helped Leow address one challenge SLA researchers face when conducting studies under an attentional framework, namely how to operationalize and measure noticing in experiments conducted in the classroom setting.

However, as in non-SLA research, the validity of think-aloud protocol as a research method is an issue of debate. The research reviewed above expressed concerns while defending its use of think-aloud protocols.

Validity of Think-aloud Protocols

As with any methodological tool, there are advantages and limitations to the use of verbal report (Gass and Mackey, 2000), including the think-aloud protocol. The advantage is that verbal report can be used to explore the participants' thinking process, which is difficult when looking only at the participants' performance in a pretest-experiment-posttest research design. However, the validity of the use of concurrent think-aloud protocols to elicit metalinguistic or nonmetalinguistic online data of learners' processes has been debated. On the one side, the widely cited Ericsson and Simmon (1993) study argued persuasively that concurrent verbal reports need not affect the processes being studied, and can be collected in ways that avoid reconstructions or interpretations on the part of participants. This argument reflected the other side's two concerns. One is the veridicality issue, i.e. whether think-aloud protocols have really

reported the participants' true and complete thinking process. While retrospective reports may be subject to the time and memory limitation between the task performed and the verbal report and therefore may allow reconstruction or interpretation, concurrent think-aloud verbal reports may also be subject to the weakness of nonveridicality due to technical or procedural issues (e.g. recording equipment's pressure on participants) in applying think-aloud protocols (Nisbett and Wilson, 1977; Olson, Duffy, and Mack, 1984). The other concern is the reactivity issue which is that the need to provide a verbal protocol, as a secondary task, may fundamentally alter the processes used in performing the primary task of interest, for example, making a choice or solving a problem. Jourdenais (2001) cautioned that "the think-aloud data collection method itself acts as an additional task which must be considered carefully when examining learner performance" (p. 373). This resonates with the concerns in psychological research as reviewed by Payne (1994), who stated, "One reason suggested for a change in processes is that the verbal protocol procedure will utilize at least some of the cognitive resources available to the respondent. Another reason may be that the need to provide a report will change what information is attended to in the stimulus; for example, information that is readily verbalizable may receive greater attention and information that is not readily verbalizable may be overshadowed" (p. 245).

Indirect evidence can be found in research in different areas. In L1 reading research, there has been evidence that think-aloud protocol may have reactivity, but on the other hand, may be used as an intervention tool in instruction. Meyers, Lytle, Palladino, Devenpeck, and Green (1990) applied think-aloud protocol to the study of tactics used by

4th and 5th-graders to facilitate reading comprehension. In the same study, Meyers et al. also reported the initial results of their follow up study designed to examine think-aloud protocol's prescriptive validity. As they reported, the patterns of moves from the initial protocols suggested useful intervention plans that resulted in an increased use of certain moves (e.g. reasoning moves); this implies that this method may have practical implications for tutoring. Another study, Afflerbach (2002), concluded that an additional value of thinking aloud is that it encourages children to spend time with their thinking and expected the conceptualization of verbal reports as aides for learning. Here thinking-aloud is supposed to result in performance difference as a result of a facilitating effect. All these studies resonate with Block's (1986) suggestion that think-aloud protocol might be a useful learning tool.

Morrison (1996) divided 20 university-level French as a second language learners in Canada into high- and low-proficiency groups and asked them to read a text individually and in pairs in a think-aloud protocol assessing the meaning of twelve underlined words. Morrison also administered a questionnaire that explored several issues including the participants' reactions to the think-aloud protocol. The positive feedback regarding the think-aloud procedure made Morrison suggest that think-aloud protocol may be used as an effective classroom tool for inference strategy teaching. The participants reported that verbalizing made them think about the meanings of the words more than they usually did and it also helped them organize their thoughts. Since think-aloud protocol is effective, we might ask whether this means some kind of reactivity in methodology.

The inconsistent findings of different researchers led to some researchers (e.g. Stratman and Hamp-Lyons 1994) taking more comprehensive views of the reactivity issue of think-aloud protocols. First, type of tasks is a matter to be considered. They reviewed that, in the extant rigorous studies of protocol reactivity, the tasks scrutinized have been more "well-defined" than "ill-defined," with the former referring to such tasks as solving mathematical problems, visual-spatial pattern problems, or decision-making problems presented in a discrete format with well-specified goals, and with the latter referring to such tasks as reading, writing, and verbal information analysis. Stratman and Hamp-Lyons emphasized that trying to extend the results of reactivity tests examing well-defined tasks to ill-defined tasks is highly problematic. Second, reactivity of think-aloud protocols may be the result of interactions between many factors. Stratman and Hamp-Lyons (1994) discussed the differential effects of the think-aloud constraint upon novices and experts and suggested that what may appear to be a difference between experts and novices may sometimes partly be an artifact produced by the interaction between the expertise a subject possesses and the constraint of giving a protocol. As Russo, Johnson and Stephens (1989) suggested, "the causes of reactivity are not general but due jointly to the demands of the task and to verbalization" (pp. 762-763).

Leow and Morgan-Short (2004) provided a review of several non-SLA studies that directly addressed the issue of reactivity but that had not been reviewed by Ericsson and Simon (1993), which suggested, in agreement with Ericsson and Simon, that verbal reports do not result in altered internal processing although extending time on task. As

the current study has reviewed, Leow and Morgan-Short (2004) is the first empirical study designed specifically to address the reactivity issue of think-loud protocol in SLA methodology, especially in attentional studies, that is, studies that "operationalize and measure the role of attention (and awareness)" (Leow and Morgan-Short, 2004, p. 36).

Leow and Morgan-Short (2004) studied the issue of reactivity of think-aloud protocols in SLA research against the background that several recent studies (e.g. Alenan, 1995; Leow, 1997, 1998a, 1998b, 2001; Rosa and O'Neill, 1999; Rott, 1999) addressed the operationalization and measurement of attention in their research by employing think-aloud protocols to gather concurrent, online data on learners' cognitive processes. As Russo et al. (1989) suggested, a useful test for reactivity can begin with output measures in carefully controlled experimentation. Leow and Morgan-Short randomly assigned 77 adult first-semester Spanish students into a think-aloud group of 38 and a nonthink-aloud group of 39 for a reading task that was followed by three assessment tasks (comprehension, intake, and controlled written production). These two groups were exposed to the same passage, pretest, and posttest assessment tasks but differed on type of condition ($\pm$ thinkaloud). The results of this study indicated thinking aloud does not affect learners' reading performance. In Leow's words, "thinking aloud while performing an L2 reading task of 384 words did not appear to have detrimental or facilitative effects on comprehension, intake, or controlled written production when compared to a nonthink-aloud performing the same task" (p. 50). Leow suggested that the predominant reading strategy (translation) revealed in the think-aloud protocols could account for the

nonsignificant difference in the amount of cognitive effort required for either reading aloud or silently, thereby reducing the potential for reactivity to play a role.

To expand on the work of Leow and Morgan-Short (2004), Bowles and Leow (2005) not only investigated the reactivity of both metalinguistic and nonmetalinguistic verbal protocol instead of just concurrent nonmetalinguistic think-aloud protocol, but also recruited 45 advanced language learners of Spanish, instead of beginners, and used a syntactic structure, instead of morphological target structure. The participants were randomly assigned into two experimental groups (metalinguistic and nonmetalinguistic) and one control group that did no verbal report. The results of the three after-reading assessment tasks, a 10-item multiple-choice comprehension task and two tasks of fill-in-the-blank written production (one for the production of the targeted structure in familiar contexts, and the other for its production in new contexts) indicated that neither type of verbalization significantly affected text comprehension or written production of old or new exemplars of the targeted structure when compared to a control group, although metalinguistic verbalization appeared to cause a significant decrease in text comprehension over nonmetalinguistic verbalization. In their study, Bowles and Leow did not attempt an explanation, as Leow and Morgan-Short (2004) did, about the non-significant difference between the control group and either experiment group who did think-aloud verbalization. Bowles and Leow (2005) seemed to put their emphasis on the similarities between those two experimental groups (e.g. a common trait is that both high and low scorers in each group reported in their protocols awareness of the targeted structure) in order to explain no significant effect from type of verbalization on the

production of target language in old and new contexts. On the significant difference between the metalinguistic group and the nonmetalinguistic group on the comprehension task, Bowles and Leow only used some comments from the metalinguistic group's think-aloud protocol to show how requesting participants' verbalization of their thoughts and justification had affected their comprehension and therefore resulted in group difference on comprehension of the text. However, participants' comments such as "it was difficult to follow the meaning of the text," as was cited by Bowles and Leow, can be used not only to explain the difference between the two experimental groups' difference but also to explain the difference between the control group and the experiment group as in Leow and Morgan-Short (2004). Furthermore, counter-intuitively, the reactivity ("difficult to follow the meaning of the text") of metalinguistic requirement could result in group difference from the non-metalinguistic think-aloud group but could not result in group difference from the control group, in the same study. Bowles and Leow (2005) did not suggest anything, as Leow and Morgan-Short (2004) did with translation as a possibility, to discuss why there was no significant difference on three assessment tasks between the control group and either of the think-aloud groups. There seems to be a lack of consistent and systematic explanation about these significant and nonsignificant differences. Therefore, although the two studies, Leow and Morgan-Short (2004) and Bowles and Leow (2005), produced somewhat consistent results about the nonreactivity of think-aloud protocols that are in agreement with the prevailing opinion in studies of other areas represented by Ericsson and Simon (1993), more empirical research is needed in SLA areas, especially a series of replication studies (along the dimensions of L2 languages, the

level of participants' L2 proficiency, etc.) so that future research may be in a better position to avoid past mistakes.

Research Questions

As aforementioned, Leow and Morgan-Short (2004) made an effort to explain the nonsignificant difference in learners' comprehension between the think-aloud and non-think-aloud groups with the finding that translation was the preferred reading strategy for many learners. They also tried to explain the assumption that the processes of translation from the L2 to the L1, silent or aloud, may not differ much in terms of required cognitive effort, thereby reducing the potential for reactivity to be an issue, if translation is also the dominant strategy employed by the non-think-aloud group. However, many studies that applied think-aloud protocols have found a variety of strategies by learners in reading (e.g. Hosenfeld, 1976). In order to check this assumption, this current study adds a fourth research question to the original three of Leow and Morgan-Short (2004) (See the Method section for rationales for these questions). The four research questions are as follows.

1. Does thinking aloud while performing a reading task have any effects (either detrimental or facilitative) on adult readers' comprehension when compared to readers not thinking aloud?

2. Does thinking aloud while performing a reading task have any effects (either detrimental or facilitative) on adult readers' intake when compared to readers not thinking aloud?
3. Does thinking aloud while performing a reading task have any effects (either detrimental or facilitative) on adult readers' controlled written production when compared to readers not thinking aloud?
4. What strategies, in addition to translation, do adult readers apply in the reading task, as are revealed in the think-aloud protocol?

Method

Participants

Participants were 30 Chinese L1 graduate students at Michigan State University. Their average length of residence (LOR) in the USA was 2.93 years, and the average of their TOEFL scores achieved before coming to MSU was 619.7 (paper test). The participants were randomly assigned into one control group and one experimental group with 15 in each. The experimental group was the group that was asked to think aloud while reading and working on the tasks. These two groups were not significantly different in terms of TOEFL or LOR (see Appendix F for more details).

Targeted Linguistic Form

The targeted linguistic form is English phrasal verbs. Although most graduate students at MSU have achieved relatively high TOEFL and/or GRE scores and show medium high proficiency of English, most of them are still weak in phrasal verbs. Consultations with some ESL instructors who teach Chinese TAs confirmed the researcher's own experience as a Chinese L1 speaker and perception of his fellows. One of the features of phrasal verbs is that learners cannot guess the meaning simply by guessing the individual parts of the phrasal verbs. However, contextual guessing (Leow and Morgan-Short, 2004, p. 44) of phrasal verbs is highly possible and guessing through context is the only way to gain comprehension if the learners do not know the phrasal verbs beforehand and do not have any reference material or people to turn to. Furthermore, without noticing the organic formation of phrasal verbs as verbs plus either separable or inseparable particles, the participants will not be able to comprehend exactly.

Reading Material

The text (see Appendix A) used in this study was an essay adapted from *Readers' Digest*: *Write Better Speak Better* (1972). In addition to the phrasal verbs originally used in the essay, some other phrasal verbs were added by a native speaker of English, who said he had tried his best to replace the original verbs with as many phrasal verbs as possible. A difference from the Leow's (2004) reading text is that this study decided not to enhance the targeted linguistic forms for two reasons: 1) The researcher believed the unenhanced text would be a better device to test the magnitude of the participants' attention to the target linguistic forms; that is to say, if the participants in the think-aloud

group took more notice of the targeted forms even in an unenhanced text than the nonthink-aloud group, it would be more reasonable to attribute the advantage in noticing to the think-aloud technique. 2) The research design could thus be simpler in that we did not have to, after having separated the participants into the think-aloud group and the non-think-aloud group, separate either group further into an enhanced group and a non-enhanced group. The text after modification, except for the phrasal verbs, was believed to be of medium-low-level difficulty for the participants.

Assessment Tasks

The rationale of designing the three assessment tasks in this study follows Leow and Morgan-Short (2004), which derived from a series of Leow's studies dating back to Leow (1997). Taking the "noticing hypothesis" of Schmidt (1990, 1993, 1994, 1995) that consciousness, in the sense of awareness of specific forms in the input at the level of noticing (conscious attention), is necessary for language learning to take place, and assuming that "if learners create a mental representation of a detected or noticed form while interacting with such a form, then their level or degree of awareness should have an impact on what they encode and later retrieve from their memory" (p. 473), Leow (1997) tried analyzing the think-aloud protocols produced by adult L2 learners of Spanish completing a problem-solving task and their immediate performance on two post-exposure assessment tasks, one recognition task and one written production task, to address the role of awareness in the human attentional system. The result did show that more awareness contributes to more recognition and more accurate written production of

targeted morphological forms. By the same rationale, the current study designed the following three tasks.

To measure participants' comprehension, a 6-item comprehension task was designed to elicit 13 pieces of information (therefore totaling 13 points in score) contained in this essay full of phrasal verbs. This comprehension task was based as much as possible on content that was related to the comprehension of phrasal verbs contained. Other general questions were also raised because one of the aims of this study was to detect the effect of the think-aloud technique upon the participants' comprehension performance. The questions were predominantly in multiple-choice or in true-or-false form. A few questions involved the participants writing a few words instead of only making choices. All the items were presented both in English and Chinese. The seventh item[2] in this task was not about the comprehension of the essay. It was a question about whether participants could realize, immediately after reading, what language structure the essay targeted. The question was placed here instead of being placed in the retrospective survey in order to avoid the possible reactive effects from the two other tasks that were placed before the retrospective survey. If the question had not been placed immediately after the reading task, it would be highly possible for the participants to realize in the process of performing the other two tasks what target language had been intended for them and any answer provided by the participants would be meaningless if the question had been in the retrospective survey (see Appendix B for details of this task).

As in Leow and Morgan-Short (2004), a multiple-choice recognition task was prepared to measure participants' intake of the targeted linguistic items, the phrasal verbs. Intake is the process of assimilating linguistic material, referring to the mental activity that mediates between input and grammars; and many factors such as comprehended input and prior knowledge of L1 and L2 are eventually important for intake (Gass and Selinker, 2001). Considering this definition, this study took Leow's (1993, 2004) definition of intake to be stored linguistic data that has been attended to by the L2 learner and may be used for immediate recognition, and intake was operationalized in this study as the participants' ability to indicate recognition of the targeted form---verb + particles ---on a multiple-choice task with the correct form and three distracters. All together 13 phrasal verbs were tested. The prompt sentences, extracted from the article in the reading task, and the choices, were all in English. The participants were also required to complete the task without going back to previous items or pages for information to avoid any potential influence of other knowledge sources on their immediate recognition of the targeted forms (see Appendix D.)

To measure participants' controlled written production of the targeted forms, a translation-and-fill-in-the-blank task comprising 29 blanks in 13 sentences was carried out. The sentences were primarily adapted from the *Longman Dictionary of Phrasal Verbs*. Although cognitively, writing develops after attention and intake, the controlled written production task in this study was administered before the multiple-choice recognition task to avoid otherwise the possible influence of the latter on the former (see Appendix C for details of this task).

The controlled written production task was carried out before the recognition task for the same principle as in Leow and Morgan-Short's research, that is, to avoid providing additional input to participants.

Testing Procedure

As Leow and Morgan-Short (2004) suggested, some guidelines for using verbal reports as summarized by Kormos (1998) from Ericson and Simon (1980, 1993) were followed as much as possible, such as asking participants to comment on their performance immediately after the completion of the task when the memory traces of the thought sequences are still fresh, providing the participants with contextual information to activate the greatest possible amount of information stored in long term memory (LTM), only requesting information related to specific problems and themes, not informing the participants of the subsequent retrospective interview (questionnaire in this study) before the completion of the task, and being invisible to the participants only taking the role of reminding participants to keep on talking while solving the given problem.

After recruiting the participants, the researcher, by manipulating names and numbers on paper, randomly assigned the participants into one control group and one experimental (think-aloud group) group. As it was difficult to get the control group participants to gather at one fixed time and place, the administration of this group's test

was divided into several sub-groups of 3 or 4 participants depending on their responses to the researcher's proposed schedule. All of the control group tests were administered in a study room in the Main Library at Michigan State University. Due to similar reasons, the experiment group participants did not do the experiment in one session, either. The difference from the control group was that an appointment was made with every participant individually because every participant was recorded individually as a result of an inability to use a language lab for recording at the same time and to avoid inter-participant influence in the process of recording.

At the time of the experiment, a package of materials containing the consent form, reading materials, assessment tasks, and the retrospective protocol stimulation sheet, was ready for each participant. The participants were told not to turn the pages until they were told to do so. Particularly, the participants were kept from knowing, at the beginning of the experiment, that there would be a retrospective protocol task, to avoid any possible reactivity on the thinking process (Kormos, 1998). Both groups were also reminded that the tasks after the reading material were both about content comprehension and about the language used. The participants in both groups were told that they might choose to read the task description in either English or Chinese because the descriptions were in both languages.

Then for the think-aloud group, the participants were told by the researcher in Chinese that they would be recorded and that there was a training session. The participants were told that this research was intended to obtain some information about

their thinking process and that they should think aloud or speak out their thoughts as naturally as they could, either in English or Chinese so long as they felt comfortable. They were also reminded that they should *not* describe or explain what they are doing *but only* verbalize the information they attend to (Ericsson, 1993). An example was offered with a simple arithmetic calculation task and a long-sentence reading task in their packet, illustrating to the participants what would be and what would not be regarded as thinking-aloud. Then the researcher asked the participants whether they had fully understood what they were expected to do. After the participants confirmed with "yes," they were also told not to worry about the time limits of the tasks. Then the experiment started: the participants put on the headsets and started reading and being recorded. In this process, the researcher managed to hide the notebook computer equipped with the recording software *Audacity*. While the participants were doing the tasks, the researcher noted as many observations as possible of the participants' language or behavior. The nonthinking-aloud group were just told to do the reading and assessment tasks as if doing exercises in normal classes.

Finally, both groups completed their retrospective report of two different versions (see Appendices E and F). The whole process for the control group was about 25-30 minutes and the process for the experimental group was about 35-40 minutes.

Choice of Language for Reporting

In addition to Nyhus (1994), which suggested that there may be a second-language threshold below which attempts to provide verbal reports in the target language will be counterproductive, Upton (1993) found that, when given a choice as to language for verbal reporting, the more advanced native-Japanese-speaking EFL participants were likely to choose to provide verbal reports on English reading comprehension tasks in English, while less proficient respondents preferred to use Japanese. Therefore, this research allowed the participants to use whatever language (either Chinese or English) they preferred in thinking-aloud, that is, to make their natural choice of language (see Table 4 in the Results section below for a report of the language chosen for reporting).

Scoring Procedure

For the comprehension and recognition tasks, one point was awarded for each correct answer, and no points for incorrect answers, for a total of 13 possible points for the comprehension task and 14 points for the recognition task. For the controlled written production task, if the participants avoided phrasal verbs in the article, no points were given. However points were given if the participants used two phrases from the article similar or close in meaning in sentences that they had not been intended for. For example, points were given for *speak up* in Question 6 and *sound off* in Question 7 in the controlled written production task. One point was awarded for the appearance of any desired base verb or particle; no point was awarded for the position of a noun or pronoun, whether correct or incorrect. There were three pronoun positions designed to be blanks as the participants would sense out phrasal verbs basing their judgment on the positions of

the pronouns if the pronoun positions were not kept blank. The total possible for the written task was 27. The overall total score for the three tasks was 54.

Transcribing and Coding the Think-aloud Protocols

The main purpose of this study was methodologically to measure reactivity through the participants' performance on the tasks after the reading task itself. Therefore, following Leow and Morgan-Short (2004), this study principally did not discuss which target linguistic forms the participants paid attention to and which not. However, in order to test whether translation does, as suggested by Leow and Morgan-Short (2004), play a role in the issue of the reactivity of the think-aloud protocol, this study coded translation in the think-aloud protocol, which covered both the reading task and the third assessment task, the multiple-choice recognition task. Translation in the other two assessment tasks was not coded because those two tasks themselves were either presented in both languages or were translation in nature.

Transformations of oral reports into written documents that eliminate features of spoken production may miss crucial interpretive resources. For instance, increased pauses, fillers, and a slowed speech rate may suggest a high processing load (Kasper, 1998). Kormos (1998) also noted that participants not mentioning something in their commentaries or reflections might suggest that they were performing a function automatically without being aware of the processes involved. Those processes may or may not be translation. Furthermore, it is very difficult to code the translation that only

silently happens in the participants' mind. The only way for this study to measure translation was to code the translation that appeared and can be literally transcribed in the think-aloud protocol by frequency, or by time, and to code it proportional to each participant's total instances or total time being recorded while thinking aloud. This study chose to count how many instances of translation (including translation of a single word, phrase, or sentence) there were in a participant's think-aloud protocol. Three types of instances in Chinese in the protocol were considered translation: a) A sentence/phrase/word in Chinese following an English sentence/phrase/word from the text with the equivalent meaning; b) A sentence/phrase/word in Chinese not following an observable English sentence/phrase/word in the protocol but traceable to an English sentence/phrase/word in the original text with the equivalent meaning; and c) A Chinese sentence/phrase/word summarizing the general meaning of a paragraph or the passage with the equivalent Chinese word for the English key word in the passage (see Table 4 for a report of the number of sentences/phrases/words coded as instances of translation.) Although a sentence is composed of phrases and words, an instance of sentence translation is not coded again as instance of phrase or word translation.

For example,

Type A translation: --- (Reading) *Make him against you...Against* 是*反对* [trans: *Against* means *objection*].

Type B translation: ---...第一个就是*知道人家怎么想的*，... [Trans: ...The first is to *know what they think...*](Neither before nor after this Chinese sentence is there any English word, phrase, or sentence in this participant's protocol that may form a relationship of translation with it. However, in the original text there is an English phrase

corresponding (also in time line) to this: *keeping in touch with what their constituency thinks*.)

Type C translation: --- (In summarization at the end of a paragraph) 就是说比较*冷静*的时候写[Trans: that is, write when you are *calm*.] (This Chinese sentence does not correspond either in meaning or in time to any sentence in the original text, but the key word *calm* in that paragraph is reflected in this Chinese sentence.)

Other Types of Data[3]

Verbal reports can, and usually do, comprise some combination of different types, that is, self-report, self-observation, and self-revelation (thinking-aloud) (Cohen, 2000). Camps (2003) also pointed out some benefits in combining concurrent and retrospective verbal reports as tools to better understand the role of attention in second language tasks. Although, as aforementioned, the design of this study was to measure reactivity through the participants' performance on the tasks after the reading and therefore did not focus on coding whether or how the participants paid attention to the targeted linguistic form, the present study did include a questionnaire that is a retrospective report and at one point a stimulated recall in order to triangulate the data in the online think-aloud protocol. In methodology, since the use of verbal reports is to obtain information that is impossible for the pretest-instruction-posttest scheme to provide (Camps, 2003), why do we not apply verbal protocol again (retrospective protocol in this study) to search for information that is not necessarily available in the after-reading task scores to detect possible

reactivity of the think-aloud technique on the thinking process? In other words, we may use verbal reports (retrospective) to study the issue of reactivity of verbal reports (concurrent think-aloud). Note, however, that this was not a typical verbal report and the participants mainly responded to rating scales (see Appendix E[5] for details).

In addition to the above types of data, this study also explored the researcher's notes of the behavior and speech of the participants before, during, and after the reading task.

Results

The data were submitted to the Statistical Package for the Social Sciences (SPSS) with the alpha level set at .05. Group statistics for the three tasks are displayed in Table 3.

Table 3. Group statistics.

	Group	N	Mean	S.D.	Std. Error Mean
Comprehension	Control	15	10.00*	2.299	.594
	Think-aloud	15	8.27	1.534	.396
Recognition	Control	15	9.80**	2.145	.554
	Think-aloud	15	8.27	2.434	.628
Controled written	Control	15	9.67	4.152	1.072
	Think-aloud	15	9.07	3.390	.875

Note: * shows significance at .05 level. **shows significance at .1 level.

First, the data was submitted to a two-tailed t-test for equality of means[4]. For Research Question 1, that is, whether thinking aloud while performing a reading task has any effects on adult readers' comprehension when compared to readers not thinking aloud, the result ($t=2.43$, $p=.022$), with the control group performing significantly better than the think-aloud group, gave a positive answer. For Research Question 2 (assessed

with the recognition task), that is, whether thinking aloud while performing a reading task has any effects on adult readers' intake when compared to readers not thinking aloud, the result (t=1.83, p=.078) did not give a positive answer at the level of .05 but gave a positive answer at the level of .1, still with the control group performing better than the think-aloud group. For the third research question, that is, whether thinking aloud while performing a reading task have any effects (either detrimental or facilitative) on adult readers' controlled written production when compared to readers not thinking aloud, the result (t=.43, p=.668) did not give a positive answer either at the .05 level or at the .1 level.

The data were also submitted to effect size testing. The effect sizes for the three tasks, in the order of the research questions were 0.89, 0.67, and 0.16, which means a large, medium, and small effect size respectively.

In summary, thinking aloud while performing a reading task seems to have detrimental effects on learners' comprehension and intake, but did not seem to affect controlled written production.

The fourth research question, that is, what other strategies in addition to translation the participants may apply in the reading task, did not include data on which statistics could be run. This study is set up in Table 5 to show the diversity of strategies the participants took in L2 reading processes.

Table 4. Amount of Translation and Choice of Reporting Language.

Protocol Number	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	Mean
Instances of translation	1	0	1	1	2	0	0	0	0	3	0	5	0	1	0	0.93
Reporting Language	C+E	E	C	C	C	E	C+E	C	E	C	C	C	C	C	E	C=9 E=4 C+E=2

C=Chinese, E=English. "C+E" means the participant alternated between Chinese and English in reporting.

Table 5. Varieties of non-translation strategies revealed in the participants' think-aloud protocols.

Strategies	Samples
Repetition	*More power to you, more power to you...*
Self-asking	*Put across*, what does that mean?
Problem assessment	(Focusing on the problem and commenting) don't understand.
Summarization	[So, the general idea is...]
Turning to larger structure for clues to understand local difficult problems	*So and so*? [not quite clear now, may be clearer after reading the whole paragraph, go on then.]
Making use of private experience	[This third point is not proper. Last week my teacher complained about this.]
Reading content word only	*Write in a **reasonable** "tone of voice."* ("reasonable" was the only word pronounced.)
Paraphrase	*Help you register*, [this is to let you] show up.
Constant understanding check	This is clear...this, not clear

Note: 1. Words in italics were those in the text. 2. Words in [ ] were translated by the researcher. 3. All the words in the Samples column were from the protocol, that is, what the participants thought aloud.

If we consider the amount of translation coded in Table 4, Table 5 should make us aware that translation was not the dominant strategy the participants took while reading in L2. In particular #27, for example, did the most translation (5) in the think-aloud group. However, her protocol included nine non-translation strategies. The percentage of translation in her strategy use was only 5/14=36%. There is no basis for us to claim

generally that translation was the dominant strategy the participants took in the process of second language reading.

Discussion

1. Is Translation Related to Reactivity?

Research Question 4, that is, what other strategies in addition to translation may the participants apply in the reading task as are revealed in the think-aloud protocol, is discussed first because the other three questions depend on this discussion. Table 4 and Table 5 show that translation is not the only strategy and not even the dominant strategy the participants applied. Therefore, translation as a strategy is not likely to be a proper explanation for the difference or similarity between the control group and the think-aloud group as Leow and Morgan-Short (2004) suggested.

One counter argument may be that the silent translation was not coded and that if silent translation had been coded, the amount of translation displayed would have increased greatly. However, it could not have been coded. Mental translation is outside of the scope of this study. In addition, mental translation is defined by Kern (1994) as the "mental reprocessing of L2 words, phrases, or sentences in L1 forms while reading L2 texts" (p. 442). It was reported by Upton and Lee-Thompson (2001) as only one of the variables involved in and influencing the L2 reading process. Lee-Thompson (2001) also

reported in his study that L2 readers use their L1 for more than just mental translation and that all of the intermediate and advanced ESL students and four of the five post-ESL students in his study also used their L1 to accomplish metalinguistic functions: making observations about the text or their reading behavior, or choosing to take some action based on the text or the reading demand.

Even if translation is the predominant strategy employed by L2 learners in L2 reading tasks, can this still be used to explain the experiment result---group differences---in this study as Leow and Morgan-Short (2004) did with their nonsignificant result? It makes no sense to give one explanation for two opposite results. However, if translation is only one of the variables that are involved in and influence the L2 reading process, we may ask another question: Is translation a factor that results in group similarity or group difference?

Based on this seemingly contradictory observation, the current study makes the following hypotheses: 1) Translation is basically one of the factors that leads to group similarity, as is explained in Leow and Morgan-Short (2004); 2) Besides translation, there are other factors that may lead to group difference; 3) Whether the thinking-aloud protocol is reactive is the result of the struggle between translation and other factors that tend to lead to group difference; and 4), since translation, the reliance on L1, decreases with the improvement of L2 proficiency, the higher the proficiency of the participants in the research that applies the think-aloud protocol is, the more likely reactivity play a role.

Bowles and Leow (2005) used advanced learners of L2 Spanish, but still no significant reactivity was displayed. This may be because those learners were still not advanced enough. They were just fifth semester students and most might not have lived for a meaningfully long time in a country with Spanish as L1, unlike the Chinese students in the current study, who had been studying English for at least 10 years and had been in the U.S.A. for an average of 2.93 years.

Then what are the other factors in this current study? The researcher's notes on the behavior and speech of the participants before, during, and after the reading task revealed that the participants complained most about not being used to thinking aloud. Some said they had never been trained to think aloud even though they were trained to do so in this study; some said thinking aloud made them unable to concentrate on the reading; one even said to think aloud is the habit of people of certain cultures but not the Chinese culture. Although it has been recognized to be important to offer training to participants before experiments, the efficacy of training deserves attention.

The language chosen for reporting may add to or reduce the influence of translation. In this study most participants in the think-aloud group chose to report in Chinese. Therefore, although as proficient learners of English they were already less reliant on L1, the existing reliance on L1, together with the drive of Chinese as the reporting language, perhaps made the think-aloud group do more translation than the control group. As Kern (1994) pointed out, if readers dwell primarily on "transformed" L1 representations rather than on the original L2 forms during much of the meaning-integration process, the

written L2 input may, in such circumstances, have little impact on the learner's acquisition of the L2 forms.

The above factors combined and reacted with each other to result in the group differences on the two research questions about comprehension and intake, that is, Research Questions 1 and 2. Some reactive factors from the think-aloud process had more influence than some leading-to-similarity factors such as translation, so that the nonthink-aloud group outperformed the think-aloud group. Why was there no significant group difference on the research question about controlled written production? Perhaps the answer still lies in translation. The controlled written production task took the form of sentence translation. When carrying out this task, the think-aloud group and the control group were really in the same situation, doing nearly the same amount of translation. Meanwhile, at such a later and deeper stage as written production in the whole process of language acquisition, the nonthink-aloud group's advantage built up at previous stages should have become weaker. Reactivity of think-aloud protocols as the result of interactions between many factors can be traced back to its studies in non-SLA areas, such as Stratman and Hamp-Lyons (1994), as reviewed previously. Differential effects of the think-aloud constraint upon novices and experts in their study, reflected in SLA, are the variable of the proficiency level of the L2 learners, which is then closely related to the amount of translation. Although we have only suggestive evidence that reactivity is the result of interactions between various factors, we can not yet rule out this possibility.

.

2. What Did the Survey (Retrospective Report, see Appendix E) Reveal?

To a certain degree, the statistics mentioned above are in agreement with the findings from the participants' retrospective reports[2] in the form of ratings in the questionnaire assigned. The think-aloud group's comment on the think-aloud protocol in the retrospective protocol is summarized in Table 6 below. And the participants' ratings

Table 6. Think-aloud group's comment on think-aloud.

Effect of think-aloud	Number of participants and percentage of whole group (15)	Key words in specific comments
No effect at all, either good or bad.	5/15, 33.3%	Only speed was lowered.
Facilitative.	4/15, 26.7%	No hurt; help activate thinking; help think.
Detrimental.	6/15, 40%	Interrupt thinking; not used to thinking aloud; affect thinking; affected by headphones.

on their own certainty about their choices and impression of the target language in the reading text when completing the three different assessment tasks (e.g., How do you rate your assurance or confidence when making the choices?) were also submitted to two-tail t-tests assuming equal variance. Although the control group's assurance ratings and impression ratings for the comprehension task and the controlled written production task were not significantly higher than those of the think-aloud group respectively, the two ratings of the control group in the multiple-choice recognition task were significantly higher than those of the think-aloud group, with $t=2.11$, $p<.05$ for certainty of choice, and $t=3.26$, $p<.005$ for impression of target language. In other words, the control group participants were more confident in their memory of the text while the think-aloud

group's confidence was lower because of the reactivity of thinking-aloud on attention and intake.

3. What Tasks Are Suitable for the Think-aloud Protocol?

The strong conceptualization of reading as cognition and the strong defense of protocol analysis as a means to investigate reading contributed to initial investigations of readers' strategies, and the last two decades have witnessed burgeoning use of protocol analysis to investigate acts of cognition, response, and reading related phenomena (Afflerbach, 2002). With greater and greater use of verbal protocol analysis in various areas of SLA, more research is necessary to review not only its validity in general but its difference in validity across various aspects of SLA or even across participants of various characteristics. For instance, Hosenfeld (1984) suggested using the think-aloud approach with students who translate and the introspective/retrospective approach with students who do not translate. From another perspective, Krings (1987), as reviewed by Kern (1994), suggested that thinking aloud is a particularly appropriate and valid way of looking at translation processes, pointing out, "since translation is, by its very nature, a linguistic process, the verbalization externalize linguistically-structured information and can normally do without an additional process of verbal encoding" (p. 166).

Payne (1994) answered yes to his own question: Are some tasks better suited to be studied using verbal protocols than other tasks are? He pointed out that, in particular, the more a task involves higher level cognitive processes that take more than a few seconds

to perform, and the more the task involves verbal types of information, the better. Someren et al. (1994) were careful to point out that the think-aloud method "is a means to validate or construct theories of cognitive processes, in particular of problem-solving" (p.9). Pressley and Afflerbach (1995) indicated, "fully automatic processes are difficult to self-report. They occur very quickly, so much so that intermediate products of processing are not heeded in short-term memory and, thus, not available for self-report. Protocol analysis is much more sensitive to processes that have not been automatized, ones that are still under conscious control" (p. 9). Although this comment is on self-report, the same situation may as well exist on self-revelation, that is, the concurrent think-aloud protocols.

We may also remember Stratman and Hamp-Lyons's (1994) differentiation between well-defined and ill-defined tasks. The current study suggests it would also be problematic to simply use the result of non-reactivity in non-SLA research (mostly with well-defined tasks) to claim or back up non-reactivity of the think-aloud protocol in SLA research, where the tasks are mostly ill-defined. More research is necessary before we may have more confidence in answering what tasks are suitable for think-aloud protocols.

One step further, based on this research and insight from the above studies, maybe we have to raise one question: is attention research in SLA suitable for the application of concurrent think-aloud protocol? Attention is really something difficult to think aloud concurrently. It should deserve the time if effort is made to review the literature in the field of SLA, and try to figure out the variables or factors that influence the issue of

validity and to provide a better reference for future studies that are interested in applying the think-aloud technique.

We may feel the need to reconsider the application of think-aloud protocol to **attention** studies in SLA when we review Ericsson and Simon (1987)'s definition of verbal protocol:

To obtain verbal reports, as new information (thoughts) enters attention, the participants should verbalize the corresponding thought or thoughts...the new incoming information is maintained in **attention** until the corresponding verbalization of it is completed (p.32, emphasis mine).

Conclusion

This study, designed to replicate Leow and Morgan-Short (2004), achieved quite different results: for this sample of participants (Chinese graduates), thinking aloud while performing an L2 reading task appeared to have detrimental effects on comprehension and intake, but no effect on controlled written production. In other words, thinking aloud[6] was reactive in this study, at least in some aspects. However, this study is not fully counter to Leow and Morgan-Short's (2004) explanation of no reactivity with translation, whether aloud or silent, being the shared predominant reading strategy between the think-aloud group and the nonthink-aloud group. However, this study further hypothesized that while translation is a factor that tends to lead to no difference between groups, it is only

one of the factors that finally determine whether the think-aloud protocol's reactivity is displayed. Therefore, this study indicates that the think-aloud protocol is not simply reactive or nonreactive. It is a matter of dynamic interaction between several factors. Future studies that plan to apply the think-aloud protocol for data eliciting need to consider characteristics of the participants such as L2 proficiency, culture, and other details so that reactivity can be reduced to the lowest degree.

This study also discussed the suitability of think-aloud protocols for attention studies in SLA. In general, this study does not object to the use of the think-aloud protocols for attention studies. However, this method can only provide some insight into the participants' thinking process. Because participants' report of their thinking process depends on whether they are *aware* of those processes, as suggested by Pressley and Afflerbach (2002), to research into attention-related topics by means of think-aloud protocols tends to be a subtle issue. Any conclusion drawn from observation of the think-aloud protocol is not firm if it is not supported by data elicited by other means. As was pointed out by Whitney and Budd (1996), although the think-aloud method can offer a fairly direct spotlight on how the contents of working memory change online during comprehension, it is like all other techniques that are used by cognitive psychologists---it is best used in conjunction with other complementary techniques.

Limitations and Future Directions of Research.

Although this study was designed to replicate Leow and Morgan-Short (2004), exact replication is impossible given that a replication study will deal with different individuals (Polio and Gass, 1997). Future replication studies may try to use participants with lower L2 proficiency. Another issue that deserves consideration is that the target linguistic form in this study was phrasal verbs, which, usually formed from very familiar words, might not work as well as those target linguistic forms in Leow's studies. More replication studies, although with different participants and instruments, will definitely contribute to a clearer picture of the validity issue of the think-aloud protocol in SLA research. One other concern about this study is that the participants in the think-aloud group were asked to think aloud not only in the process of reading but also in the process of completing the assessment tasks. Although this study was only replicating Leow and Morgan-Short (2004) and not considering making major modifications in research design, it should be meaningful for future research to consider asking participants only to think aloud while engaged in the reading task itself because the intended purpose of Leow and Morgan-Short (2004) had been to investigate empirically the issue of reactivity of thinking-aloud on the reading process but not on assessment tasks. Future research may also investigate the role that participants' different reporting languages may play in the issue of reactivity of the think-aloud protocols. Due to the limits of focus and time frame for this study, some meaningful questions were not asked, such as those about why the participants chose to report in one language rather than the other, and what they feel about using a certain language for reporting. As Leow and Morgan-Short (2004) suggested, the issues of reactivity of think-aloud protocols are clearly fruitful areas of investigation in SLA research methodology.

Notes

1. Refer to Gass and Mackey (2000) for a complete list of references included in the list of SLA studies that applied think-aloud protocols.

2. The question of what target language the participants realized the reading material had been intended for them to learn (the last question in the comprehension task) did not produce clear-cut responses. Only three participants roughly figured out the language purpose of the reading material was phrasal verbs. The participants might not really have "noticed" it, or might have misunderstood this question.

3. On the two retrospective questions about whether the participants could remember the existence of the two phrasal verbs *tip off* and *tell off*, the two groups reported equal or very close cases of remembering. This was not discussed in the study because these two phrases were only two out of nearly 15 phrasal verbs for the participants. The researcher hopes it may be useful for other researchers.

4. As this study only recruited a small size of participants, to be conservative, the data were submitted to nonparametric t-tests (Mann Whitney U). For Research Question 1, the result t=56.5, p=.018 showed a significant difference in comprehension performance between the control group and the think-aloud group with the former group outperforming the latter. For Research Question 2, the result (t=72.5, p=.092) was not significant at the .05 level but at the .10 level a significant difference in immediate multiple-choice recognition performance between the control group and the think-aloud group still with the former outperforming the latter. For Research Question 3, the result (t=111.50, p=.967) showed no significant difference between the two groups in the controlled written production either at the .05 level or at the .1 level. This result totally agrees with the result of the independent samples t-test for equality of means.

5. The survey for retrospective reports has two different versions for the think-aloud group and the nonthink-aloud group. The version for the nonthink-aloud group is the whole set of the questions in the version for the think-aloud group without the first question which asks how the think-aloud participants perceive the technique used upon them.

6. The think-aloud here is non-metalinguistic, as this study and Leow and Morgan-Short (2004) were intended for. The significant difference between the metalinguistic and non-metalinguistic groups in Bowles and Leow (2005) reminded this study of whether the participants in this study were really doing non-metalinguistic instead of metalinguistic think-aloud protocols. After a post hoc rough observation of the participants' protocols, it can be concluded that the participants in this study were indeed doing non-metalinguistic thinking-aloud. We see very few instances of metalinguistic think-aloud protocol such as "I think this is right because..." Only one participant can be observed to have made some metalinguistic think-aloud protocol such as "This is right. I completely agree to this...My

teacher told me last week not to write too abstract or simple things but to use rich vocabulary...I know what *boil down* means but why it is used this way, I don't know..." It is necessary for future research using think-aloud protocols to consider the different effects of metalinguistic and non-metalinguistic protocols. Those methodological studies that investigate the reactivity issues of metalinguistic and non-metalinguistic protocols may also need to examine the participants' protocols to see to what degree the participants have performed according to the researchers' requirement.

References

Afflerbach, P. (2002). Verbal reports and protocol analysis. In Kamil, M. L., Mosenthal, P. B., Pearson, P. D. and Barr, R. *Methods of literacy research*. Lawrence Erlbaum Associates.

Anderson, N. (1989). *Reading comprehension tests versus academic reading: What are Second Language readers doing?* Unpublished doctoral dissertation, University of Texas at Austin.

Block, E. (1986). The comprehension strategies of second language readers. *TESOL Quarterly, 20*, 463-494.

Bowles, M. A. and Leow, R. P. (2005). Reactivity and type of verbal report in SLA research methodology: Explaining the scope of investigation. *Studies in Second Language Acquisition, 27,* 415-440.

Camps, J. (2003). Concurrent and retrospective berbal reports as tools to better understand the role of attention in second language tasks. *International Journal of Applied Linguistics, Vol. 13, No. 2,* 201-221.

Cavalcanti, M. (1987). Investigating FL reading performance through pause protocols. In C. Færch and G. Kasper (Eds.), *Introspection in second language research* (pp.230-250).

Cohen, A. D. (2000). Exploring strategies in Test-taking: Fine-tuning verbal reports from respondents. In G. Ekbatani and H. Pierson (Eds.), *Learner-directed assessment in ESL* (p127-150). Mahwah, NJ: Laurence Erlbaum.

Courtney, R. (1983). *Longman Dictionary of Phrasal Verbs*. NY: Longman.

Ericsson, K., and Simon, H. (1987). Verbal reports on thinking. In C. Farch and G. Kasper (Eds.), *Introspection in second language research* (pp. 24-53). Clevedon: Multilingual Matters.

Ericsson, K., and Simon, H. (1993). *Protocol analysis: Verbal reports as data (Rev. ed.)*. Cambridge, MA: MIT Press.

Fraser, C. (1999). Lexical processing strategy use and vocabulary learning through reading. *Studies in Second Language Acquisition, 21*, 225–241

Gass, S. and Selinker, L. (2001). *Second language acquisition: An introductory course.* Mahwah, NJ: Laurence Erlbaum.

Gass, S. and Mackey, A. (2000). *Stimulated recall methodology in second language research.* Mahwah, NJ: Erlbaum.

Hosenfeld, C. (1976). Learning about learning: Discovering our students' strategies. *Foreign Language Annals, 9,* 117-129.

--- (1977). A preliminary investigation of the reading strategies of successful and nonsuccessful second language learner. System 5: 2. 110-123.
--- (1979). Cindy: A learner in today's foreign language classroom. In W. Borne (Ed.), *The Foreign language learner in today's classroom environment.* Northeast Conference on the Teaching of Foreign Languages.

---(1984). Case studies of ninth grade readers. In Alderson J.C. (Ed.), *Reading in a foreign language.* Longman: London and New York.

Jourdenais, R. (2001). Protocol analysis and SLA. In P. Robinson (Ed.), *Cognition and second language acquisition* (pp. 354-375). New York: Cambridge Universtiy Press.

Kasper, G. (1998). Analysing Verbal Protocols. *TESOL Quarterly, 1998, 32, 2,* 358-362

Kern, R.G. (1994). The role of mental translation in second language reading. *Studies in Second Language Acquisition, 16,* 441-461.

Kormos, Judit (1998) The use of verbal Reports in L2 research -verbal reports in L2 speech production research. *TESOL Quarterly, 32, 2, 1998,* 353-358

Leow, R.P. (1997). Attention, awareness, and foreign language behavior. *Language Learning 47*, 467-505.

Leow, R.P. (1998a). The effects of amount and type of exposure on adult learners' L2 development in SLA. *Modern Language Journal 82*, 49-68.

Leow, R.P. (1998b). Toward operationalizing the process of attention in second language acquisition: evidence for Tomlin and Villa's (1994) fine-grained analysis of attention. *Applied Psycholinguistics 19*, 133-159.

Leow, R.P. (2001). Do learners notice enhanced forms while interacting with the L2? An online and offline study of the role of written input enhancement in L2 reading. *Hispania 84,* 496-509.

Leow, R. P. and 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.

Meyers, J., Lytle, S., Palladino, D., Devenpeck, G. and Green, M. (1990). Think-aloud protocol Analysis: An investigation of reading comprehension strategies in fourth- and fifth-grade students. *Journal of Psychoeducational Assessment, 1990, 8, 2, June*, 112-127.

Morrison, L. (1996). Talking about Words: A study of French as a second language learners' lexical nferencing procedures. *The Canadian Modern Language Review/La Revue canadienne des langues vivantes, 1996, 53, 1, Oct*, 41-75

Nisbett. R. E., and Wilson, T. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84,* 231-259.

Nyhus, S. E. (1994). *Attitudes of non-native speakers of English toward the use of verbal report to elicit their reading comprehension strategies*. Unpublished master's thesis, University of Minnesota, Minneapolis.

Olson, G., Duffy, S. A., and Mack, R. L. (1984). Thinking-out-loud as a method for studying real-time comprehension processes. In D. E. Kieras and M. A. Just (Eds.), *New methods in reading comprehension research* (pp. 253-286). Mahwah, NJ: Erlbaum.

Payne, John W. (1994). Thinking aloud: Insights into information processing, *Psychological Science, 5, 5, September 1994*, 241-248 (review article)

Polio, C., and Gass, S. (1997). Replication and reporting: A commentary. *Studies in Second Language Acquisition, 19,* 499-508.

Pressley, M., and Afflerbach, P. (1995). *Verbal protocols of reading: The nature of constructively responsive reading.* Mahwah, NJ: Erlbaum.

Russo, J., Johnson, E., and Stephens, D. (1989). The validity of verbal protocols. *Memory and Cognition, 17 (6)*, 759-769.

Shavelson, R., Webb, N., and Burstein, L. (1986). Measurement of teaching. In M. Wittrock (Ed.), *Handbook of research on teaching* (pp. 50-91). New York: MacMillan.

Smith, Michael W. (1991). Constructing meaning from text: an analysis of ninth-grade reader responses. *The Journal of Educational Research, 1991, 84, 5, May-June*, 263-271.

Someren, M. van, Barnard, Y., and Sandberg, J. (1994). *The think aloud method: A practical guide to medelling cognitive processes.* London: Academic Press.

Stratman, J. F., and Hamp-Lyons, L. (1994). Reactivity in concurrent think-aloud protocols. In P. Smagorinsky (Ed.), *Speaking about writing: Reflections on research methodology* (pp. 89-112). London: Sage.

Upton, Thomas A and Lee-Thompson, Li-Chun. (2001). The role of the first language in second language reading. *Studies in Second Language Acquisition, 2001, 23, 4, Dec*, 469-495

Whitney, P. and Budd, D. (1996). Think-aloud protocols and the study of comprehension. *Discourse Processes, 1996, 21, 3, May-June*, 341-351

Appendices

Appendix A **THE READING TEXT**

Please read the following article. As you read the article, THINK ALOUD your thoughts as naturally as you can. You may also make any mark necessary on the article while you are reading. When you are finished, please turn the page and complete the following tasks. You can now turn on the recorder and start reading. (请阅读下面短文.阅读的同时要将自己的思维"自言自语"出来.你可以在文章上面作任何必要的记号. 读完以后, 翻到下一页, 完成后面的练习. 注意, 在做练习时，你是不能再返回来参看这篇文章的。你现在可以打开录音机并开始阅读了.)

OVER a good many years working at a farm magazine, I have run into all kinds of letters. Some have spoken favorably of us, more haven't, and that's what I have been preparing for. Those who disagree with the editor's views, or with something else showing up in the magazine, are more likely to write and speak up. They have as much right to their opinion, and to sound off, as the editors have. And that's fine. It's great that we live in a country where it's that way.

But if you do write a letter of disagreement, I'll tip you off as to how you can hit the editor hardest:

1. Write in a reasonable "tone of voice," even if you're boiling mad. If you're writing just to tell off the so-and-so, go ahead if it makes you feel better. The letters that have some thought to put across, and that do it in a calm, unshrill way, are the ones that break through editor's hides and really "get to" him. A sincere letter like this can have more impact on us than you might ever guess---whether or not it winds up in print in the Letters column.
2. Disagree all you want with a statement, an idea, or a point of view, but don't attack the editor's motives. You only get his defenses up when you lash out at his motives; and you certainly won't win him over that way.
3. Make the letter reasonably brief. If you can boil a long story down to a few sentences, it still has the same meaning. The three-page, single-spaced kind is just too much to expect anyone even to look at in the hectic days common to any lively editorial office. It will get more attention if when the editor picks it up he sees it is of moderate length. Besides, in such length you'll probably make your point more effectively anyhow.

Editors look forward to mail from readers. The more, the better. That's one of their ways of keeping in touch with what their constituency thinks. Besides, there are some mighty good ideas in that mail, some of the best that editors come across anywhere. There's stimulus there, too, and we cry for it daily.

So please write, even oftener---we love it. All I'm trying to do here is help you "register" when your letter rushes in, and I assume that's what you want.

More power to you.

Carroll Streeter

Appendix B

COMPREHENSION TASK (阅读理解)

Now see if you can answer the following questions based on the article you have just read. Answer either in Chinese or English when you have to write.. Answer every question before you move to the next. Do NOT leave one not answered and come back to it later. (请回答以下有关问题. 需写字时使用中英文皆可. 一个一个地回答, 不要空题等回头处理. 答题时, 请不要回头翻看原文.)

KEEP THIINKING ALOUD!
不要忘记 "自言自语"!

1. 请给本文取一标题. (Please give a title to the article)

_____.

2. 判断正误, 在括号内用√表示正确, 用×表示错误. (True or false judgment. Please put in the brackets a "√" for "correct" and a "×" for "wrong".)

( ) 2.1 本文作者没料到收到的赞扬信比批评信多.
(The author of this article did not expect to have received more commendatory letters than those that are not.)

( ) 2.2 文中说,多数情况下,与编辑持不同意见者会在信中委婉表达而不是直率表达.
(According to this article, people who disagree with the editor's views are more likely to express their disagreement in a roundabout way rather than outspokenly.)

( ) 2.3 文章曾提到, 读者与编辑有同等权利表达意见.
(This article ever mentioned that readers have as much right to express their opinion as the editors have.)

( ) 2.4 作者认为读者攻击编辑的动机会使编辑产生防卫心理.
(The author believes attacking an editor's motives can make him guard against you.)

( ) 2.5 根据这篇文章,给编辑的信, 比较合适的长度是单行距三页纸.
(According to this article, the proper size of a letter to an editor is three pages single-spaced.)

3. 以下表述当中, 哪些是本文作者给出的建议? 请将句子前面的序号圈起.
(Which of the following are the pieces of advice offered by the author? Please circle the serial numbers in front of the sentences.)

1) 多表达一致意见.
(Write more letters of agreement.)
2) 要求编辑解释动机
(Ask the editor to explain his/her motives.)
3) 平和地表达意见
(Express your opinions in a calm way.)
4) 将信直接送到编辑办公室.
(Take your letter directly to the editor's office.)
5) 写信要简明.
(Write reasonably brief letters.)

4. 文中提到编辑收到读者来信有三种益处. 请写在下面横线上. (According to the article, there are three benefits for an editor to keep correspondence with readers. Please list them on the lines below.)

1) _____.
2) _____.
3) _____.

5. 本文作者是什么身份? 请圈出序号. (What is the author of this article? Please circle the number.)

1) 编辑 (An editor.)
2) 读者 (A reader.)

6. 请选出最合适的作为标题, 请圈出序号.(Please choose the best out of the following as the title of this article by circling the serial number.).

1) 写信应该写什么(What to Write in a Letter)
2) 什么样的信受欢迎(What Kind of Letters Are Welcome)
3) 怎样打动编辑(How to Hit the Editor Hardest)
4) 编辑怎样鼓励读者写信 (How Editors Encourage Readers to Write)

下面一个问题不是阅读理解题，但请努力回答．（The question below is not a comprehension question. But please try your best to answer it.）

7. 现在，从英语学习的角度，你能看出或猜出本篇文章想让你学习什么语言(语法)点吗？ (Now, at the angle of English language learning, could you tell or guess what language or linguistic point this article is trying to present to you?)

_____.

Appendix C

FILL-IN-THE-BLANK CONTROLLED WRITTEN PRODUCTION TASK

请根据下面各题中中文句子的意思, 在英文句中每空填上一词, 使英文句意完整. 尽量使用你刚读过的文章中的词语（句子不一定是原句）. 但是，不要回原文查找. （Fill in one word for each blank and make the English sentences complete in meaning according to the Chinese sentences in each pair. Try to use what you have just learnt from the article you just read. But please do NOT turn back to the article for answers. ）

继续“自言自语”！
DON'T FORGET TO KEEP THINKING ALOUD!

1. 他从毕业就在那家公司干。
He's been _____ _____ that company since graduation.

2. 踏入社会，你会遇到各式各样的人。
Stepping society, you may _____ _____ all kinds of people.

3. 在投票权问题上，两名委员会成员选择了与主席不同的立场.
Two of the committee members chose to _____ _____ the chairman on the question of voting rights.

4. 光线明亮的时候，她的白发就显了出来.
Her grey hair _____ _____ in the bright light.

5. 如果你认为那是不公平的，为什么不大胆说出来呢?
If you thought that wasn't fair, why didn't you _____ _____ ?

6. 你和编辑一样有自由表达意见的权利.
You have as much right to _____ _____ as the editors do.

7. <<新市场报>>的记者给了我一些大赛马会的建议; 要不然, 我不会赢这么多钱.
The Newmarket reporter _____ _____ _____ about the big horse race; otherwise, I would not have won so much money.

8. 因为上课又迟到, 老师批他了.
The teacher _____ _____ _____ for being late for class again.

9. 起初很难突破他们的防线.
It was very difficult at first to _____ _____ their defense.

10. 那位旅客坐错了火车, 结果到了一个小山村.
That traveler took the wrong train and _____ _____ at a mountain village.

11. 吉姆突然进攻来敌, 把他们彻底打败.
Jim _____ _____ at his attackers and beat them thoroughly.

12. 那样说话你是不会把他争取过来的.
Speaking that way you won't be able to _____ _____ _____.

13. 我碰巧在抽屉最里面找到了这张老照片.
I _____ _____ this old photograph in the back of the drawer.

Appendix D MULTIPLE-CHOICE RECOGNITION TASK

以下各题中的句子都是从你刚读过的文章中抽取出来的,但是有关词语已被空出. 请从 ABCD 四个选项中识别出原文使用的词语, 将选项前的字母填入前面括号中. **请按顺序做题**. (The following sentences are extracted from the article you read but the related words have been taken place of by blanks. Please recognize, of ABCD, the one that is used in the original article and put the correspondent letter in the brackets in front. **(Please do NOT skip a number and then return to it.)**

继续"自言自语"!
DON'T FORGET TO KEEP THINKING ALOUD!

(1. 2.) 1. Over a good many years __(1)__ a farm magazine, I have __(2)__ all kinds of letters.
(1) A. working out B. working in C. working at D. working on
(2) A. run out of B. run into C. run against D. run for

( ) 3. Some have been commendatory, more haven't, and that's what I've been _____.
A. preparing about B. preparing against
C. preparing for D. preparing of

( ) 4. Those who disagree with the editor's views or with something else _____ in the magazine, are more likely to write and say so.
A. showing off B. showing around
C. showing upon D. showing up

( ) 5. They have as much right to their opinion, and to _____, as the editors have.
A. sound out B. sound off C. sound up D. sound through

( ) 6. But if you do write a letter of disagreement, I'll _____as to how you can hit the editor hardest.

A. tip up you B. tip you off
C. tip you up D. tip off you

( ) 7. If you're writing just to _____ the so-and-so, go ahead if it makes you feel better.
A. tell about B. tell down C. tell off D. tell after

(8. 9.) *. The letters that have some thought to_____(8)_____, and that do it in a calm, unshrill way, are the ones that _____(9)_____editor's hides and really "get to" him.
(8) A. put across B. put through C. put out D. put in
(9) A. break up B. break through C. break into D. break down

( ) 10. A sincere letter like this can have more impact on us than you might ever guess---whether or not it _____ in print in the Letters column.
A. winds down B. winds away C. winds up D. winds off

(11. 12. 13.) *. Don't attack the editor's motives. You only ____(11)____ when you __(12)__ his motives; and you certainly won't __(13)____ that way.
(11) A. get his defenses on B. get his defenses up
C. get off his defenses D. get out his defenses
(12) A. lash out at B. lash off to C. lash out upon D. lash off into
(13) A. win upon him B. win him in C. win over him D. win him over

( ) 14. Besides, there are some mighty good ideas in that mail, some of the best that editors _____ anywhere.
A. come into B. come across C. come on D. come at

( ) 15. All I'm trying to do here is help you "register" when your letter _____, and I assume that's what you want.
A. rushes off B. rushes out C. rushes at D. rushes in

Appendix E 回顾报告(Retrospective Report)

请按以下提示回顾自己的思维过程中某些方面。(Please answer the following questions.)

1. "自言自语"组:
你认为"自言自语"这种要求影响你的阅读理解和解题了吗？如果有，是有益还是有害？请解释。
答：A. 无任何影响，不论是益处还是害处。
B. 有益处，益处等级（由低至高）：1 2 3 4 5 6 7
因为：

C. 有害处，害处等级（由低至高）：1 2 3 4 5 6 7
因为：

2．请看阅读理解练习题。在做阅读理解题时，做的有把握吗？是很清楚地根据对文章的理解和记忆，还是根据自己常识判断和做题技巧？
答： 把握程度： 1 2 3 4 5 6 7
根据理解和记忆： 1 2 3 4 5 6 7
根据常识和解题技巧：1 2 3 4 5 6 7
补充说明：

3. 请看翻译填空题，讲讲当时的感觉．知道是填短语动词吗？填的时候对所填内容有把握吗？对原文的短语动词的印象对你有帮助吗？
答： 知道程度：1 2 3 4 5 6 7
把握程度：1 2 3 4 5 6 7
有帮助： 1 2 3 4 5 6 7
补充说明：

4. 请看选择题。选择时有把握吗？当时对文章中的有关词语有印象吗？
把握程度：1 2 3 4 5 6 7
印象程度：1 2 3 4 5 6 7
补充说明：

5. 记得原文中有 tip (you) off 这个词语吗？ 记得[ ]； 不记得[ ]

6. 记得原文中有 tell off 这个词语吗？ 记得[ ]； 不记得[ ]

实验结束， 万分感谢！

Appendix F THE PARTICIPANTS' LOR AND TOEFL SCORES

Table 7. Length of Residence (Years)

Participant #	Control Group	Participant #	Think-aloud Group
1	1.5	16	1.5
2	1.5	17	1.5
3	1.5	18	1.5
4	1.5	19	1.5
5	1.5	20	1.5
6	2	21	1.5
7	0.5	22	0.5
8	1.5	23	1.5
9	1.5	24	1.5
10	1.5	25	5.5
11	1.5	26	0.5
12	0.5	27	1.5
13	0.5	28	2.5
14	0.5	29	2
15	0.5	30	1.5

Table 8. TOEFL Scores

Participant #	Control Group	Participant #	Think-aloud Group
1	633	16	612
2	637	17	627
3	580	18	620
4	607	19	627
5	610	20	620
6	625	21	617
7	600	22	657
8	600	23	610
9	610	24	620
10	617	25	615
11	640	26	620
12	610	27	641
13	597	28	637
14	623	29	620
15	637	30	623

MICHIGAN STATE UNIVERSITY LIBRARIES
3 1293 02736 6776