# Think-Aloud Protocols and Type of Reading Task: The Issue of Reactivity in L2 Reading Research

**Mami Yoshida**
**Kyoto University of Foreign Studies**

## 1 Introduction

"Protocol analysis" or "think-alouds" or have been extensively employed in the fields of psychology and cognitive science as a verbal-report method of producing concurrent verbalization. Think-alouds require participants to tell researchers what they are thinking and doing while performing a task. The participants are usually instructed to keep thinking aloud, acting as if they are alone in the room speaking to themselves. Think-aloud protocols are tape- and/or video-recorded and then transcribed for content analysis. The think-aloud protocols are, in many cases, coded for specific categories which have previously been developed by the researcher.

In the past few decades, think-aloud protocols have been used in SLA research to observe the cognitive processes involved in the use and acquisition of language. The major SLA areas where think-alouds have been extensively utilized are Reading, Writing, and Testing. Language acquisition, discourse research, and research on attention and awareness have been also major areas which have recently benefited from this methodology. Think-aloud protocols have provided language acquisition researchers with information as to the types of strategies employed by learners when interacting with L2 tasks, what types of input induce most noticing, what types of process can be predicted by particular type of verbal reports, and so on.

L2 reading research has also relied upon think-aloud protocols to analyze reading processes, because reading is normally a silent, hidden process, researchers cannot determine what is happening by simple observation or by product-based assessment. Asking readers to provide verbal reports or protocols is the most direct way to access this process. Other methodologies to investigate readers reading processes are, strategy questionnaires, interviews, eye-movement indices, and oral reading. Because think-aloud protocols provide concurrently gathered data of verbal products, on the other hand, it enables researchers to examine in-depth and compare reading processes in particular reading conditions, such as L1 and L2 reading (Block, 1992), test taking (Alderson, 1990), L1 and L2 narrative reading (Zwaan & Brown, 1996), mental translation processes (Kern, 1994), and task conditions (Horiba, 2000).

Recent studies (Leow & Morgan-Short, 2004; Bowles & Leow, 2005) have empirically addressed the issue of reactivity on L2 acquisition – the act of thinking aloud potentially triggering changes in learners' cognitive processes while performing the task. However, the reactivity effects of thinking-aloud on L2 reading comprehension and processes have not been fully investigated, especially in relation with the type of task readers engage in while reading. The purpose of this study, therefore, is to investigate the effects of thinking-aloudon L2 reading in terms of reading time, comprehension, and interaction with while-reading tasks.

## 2 Background

### *2.1 Categorization of Think-aloud Protocol*

Think-alouds can be categorized as retrospective or concurrent. In retrospective think-alouds, participants are required, after performing a task to recall what they were thinking while they were involved in the process of completing that task. In concurrent think-alouds, on the other hand, they are

asked to say out loud what they are thinking during the actual process of completing the task.

Ericcson and Simon ( 1984, 1993 ) also distinguish between reports that require subjects to verbalize their thoughts per se and those that require subjects to verbalize additional information such as explanations and justifications. Bolwes and Leow (2005) refer to the former verbalization as non-metalinguistic and the latter verbalization as metalinguistic. It has been argued that that metalinguistic verbalization potentially promotes "dual processing" (i.e. engaging in a learning task and reporting their thoughts at the same time (cf. Ellis, 2001).

### *2.2 Benefits of Think-aloud Protocol*

One of the major benefits of think-aloud protocols is that there are no processing-reporting interval effects in think-aloud protocols; readers can report their thoughts while simultaneously being involved in the target task. Think-aloud protocol differs from retrospective verbalization, such as interviews and retrospection in its use of concurrent verbalization. As verbal reports are collected upon completion of a task as learners are prompted to think back upon and report the processes and thoughts they went through in retrospection, there are naturally latency effects in what they remember, while think-aloud protocol elicits information most recently attended by the learners.

Afflerback (2000) also pointed out, as benefits of the methodology, that think-aloud protocol yields detailed descriptions of task-induced reader behaviors and complexity in reader's thoughts and that it also permits the effect of affective states on reader-text interaction.

### *2.3 Issues of Controversy about Think-aloud Protocol*

Although many advantages to think-aloud protocol have been pointed out, there are also some concerns about this methodology expressed by cognitive psychologists and L2 researchers. The criticisms to be discussed below are concerned with two important issues: the veridicality and reactivity of verbal reports. One of the potential factors which will affect the veridicality of verbal-report data is automation of processes. It has been argued that verbal reports cannot provide a full picture of processing, especially when the text being read is so easy that reading activities are automatic and inaccessible to verbalization. It has been argued that the processes that can be observed by thinking-aloud are limited to conscious processes and automatized processes that readers do not attend to cannot be reported (cf. Lyons, 1986). As a result, the elicited data are incomplete reports which are quantitatively poor reflections of cognitive processing.

Another concern is individual differences in verbal facility. Individuals with normally developed language skills will differ in their tendency to produce verbal reports, perception about the task, and performance on the task, due to differences in gender, personality, and previous experience. Faulty reporting is also another potential source of invalidity of verbal reports. Informants know more than they can tell, while they tell more than they can know (Nisbett & Wilson, 1977).

Finally, the reactivity of the verbal reports, which is the main concern of this paper, is another controversial issue. It has been pointed out that thinking-aloud might alter reading processes because thinking-aloud differs from normal silent reading. Frequent interruptions, or heavy cognitive load as sources of the possible alternation and disruption of the processes have also been claimed to result in incomplete reporting (Selinger, 1983; Stratman & Hamp-Lyons, 1994).

### *2.4 Findings about the Reactivity Effects in L2 Research*

In spite of critiques about the reactivity issue of verbal reports, namely that the task of generating concurrent verbal reports itself may change the nature of actual cognitive processes, many studies in L1 research have found no reactive effects of the think-alouds and support Ericsson and Simon's (1993) assumption that thinking aloud does not lead to a reliable change in the cognitive process as reflected in accuracy of response. However, Leow and Morgan-Short (2004) questioned whether the assumption of nonreactivity is applicable to tasks in SLA and provided empirical data showing no reactivity on text comprehension, intake, and production with beginners. Bowles and Leow (2005) continued addressing the reactivity issue by investigating the differential effects of type of verbalization (non-metalinguistic

and metalinguistic) with more advanced language learners. They found that neither type of think-aloud protocols caused reactivity but found metalinguistic verbalization appeared to cause a decrease in text comprehension.

Subsequent studies have shown mixed results. Bowles (2008) showed that metalinguistic verbalization lengthens the time participants need to complete the task and seems to hinder item learning but concluded that simply thinking aloud (i.e. non-met linguistic) does not alter the basic underlying processes the study was set to investigate because either type of verbalization significantly affected participants' ability to produce novel exemplars of the targeted structure and no interaction was found between verbalization and feedback. Sachs and Suh (2008) which investigated the effect of textual enhancement of recasts on learning of a targeted form, indicated reactivity effect of thinking-aloud on text completion task but did not conclude that thinking aloud had affected the learners'improvement from pre- to post- task. Although complex pictures of the interaction between external conditions and reactivity have been shown, these studies basically support the position that reactivity does not play a significant role in L2 learners' subsequent learning effects.

On the other hand, there have also been studies which explicitly indicated the reactivity effects of think-alouds on L2 learning performance. Sachs and Polio (2007) investigated the differential effects of reformulations versus written error correction on awareness and L2 grammatical accuracy in advanced ESL learners and showed that the non-think aloud group which worked on the reformulations outperformed the think-aloud group. Sanz, Lin, Lado, Bowden, and Stafford (to appear) found positive reactivity on the development of a non-primary language. In their study, participants received a computerized treatment including a grammar lesson, practice and feedback, reactivity effect was found on posttest performance in terms of latency but no reactivity effect on accuracy. In their second experiment where no grammar instruction was given, facilitating reactivity was found. It was suggested that reactivity depends not only on the task and on the nature of the assessment tool, but also on the nature of the dependent variables. As Ericsson and Simon (1993) pointed out, these studies suggest that reactivity depends not only on the instructions given, but also on the characteristics of the task.

### *2.5 Reactivity on Reading*

Although the reactivity issue has rarely been a main theme of reading research by itself, the possible effects caused by thinking-aloud have been considered and it has been often pre-checked whether thinking-aloud impedes text comprehension before think-aloud protocols are analyzed. The assumption that the thinking-aloud does not intrude significantly on the comprehension processes itself has been generally assumed and supported by empirical data such as Fletcher (1986) which found no detrimental effects of thinking-aloud on L1 reading comprehension compared to silent reading and Horiba (1990) who found that thinking-aloud neither distract nor impede L1 and L2 readers' text comprehension.

Baumann, Seifert-Kessell, and Jones (1992) is one of the few researchers have directly questioned the reactivity issue in reading research. They investigated whether thinking-aloud affected L1 comprehension monitoring abilities in children while they read a story and found positive effects of think-alouds. Although the issue of reactivity was not explicitly focused, other previous studies also found positive effects of thinkin-aloud with young readers (Chi, de Leeuw, Chiu, & La Vancher, 1994; Coté, Goldman, & Saul, 1998). Loxterman, Beck, and McKeown (1994) also reported of the benefits of thinking-aloud especially when children read an incoherent text.

Thus, it may be possible to assume that thinking-aloud does not affect simple text comprehension which requires only receptive information processing skills, but can be reactive when extra cognitive demand is required, such as monitoring, providing specific information, or finding gaps in coherence of the text. Furthermore, unskilled readers such as young readers can benefit from thinking-aloud when they read a difficult text or engage in a complex task. Therefore, it may be possible to assume that reading in a second language and engaging in while-reading tasks may cause a cognitively demanding enough task to be affected by thinking-aloud negatively or positively.

Motivated by the speculations made by Bowles and Leow (2005) that reactivity varies according to task type, text variables, individual differences, and mixed results about reactivity in relation with task effects shown in the subsequent studies, the present study will investigate the reactivity effects of

thinking-aloud on L2 reading in relation with the type of task which readers engage in while reading. In order to do so, the following sub questions will be answered:

1) Do the effects of think-alouds interact with differences in while-reading tasks (i.e. outlining, answering embedded questions and reading-only) on text comprehension in terms of the number of recalled ideas?
2) Do think-alouds facilitate or inhibit text comprehension in terms of the amount of information recalled (immediately and a week later)?
3) Do think-alouds increase the amount of time on task?

## 3 Method

### *3.1 Participants*

Sixty-four participants enrolled in a four-year university in western Japan participated in the study. They are all English majors; their language proficiency level was considered to range from intermediate to upper intermediate levels (56.6 points on average, out of 80 with scores ranging from 75 to 33 on the reading section of the Michigan Placement Test). Their educational background was homogeneous and none had spent more than one year in a foreign country.

Participants were randomly assigned to a think-aloud or a non-think-aloud group. An independent *t*-test showed no statistically significant difference in scores of English reading ability (Michigan Placement Test: Form C) between the think-aloud group and the non-think aloud group ($p < .01$). There were 31 participants in the think-aloud group and 33 participants the non-think-aloud group.

### *3.2 Materials*

An expository text was used in the experiments. The passage was selected from an ESL reading text, *More Reading Power* (Mikulecky & Jeffries, 1996), in which it was part of a series of rapid reading activities. The expository passage, which was about deforestation, contained 488 words, 40 sentences, and had a Flesch–Kincaid Readability Index of 6.4 and a Flesch Reading Ease rating of 71.2.

Three sets of reading materials were prepared. The set prepared for the control condition (i.e., the read-only group) included a passage printed on the first page and a sheet for recall after the passage. For the group of participants who answered embedded questions during reading, the entire reading passage and all questions were printed on the left side of the same page. Furthermore, five questions were written for the nine paragraphs with one question following every two paragraphs. For the outlining condition, the entire reading passage, instructions, and examples of outlining were printed on the same page. Participants in this condition received the passage with a brief explanation of how to outline and an example of an outline in English on the right hand side. It was pointed out that the two tasks differ from each other in terms of cognitive demand by Yoshida (2007) showing that outlining takes more time and encourages the use of higher-level reading processes more than embedded questions.

### *3.3 Procedure*

First, general instructions for the experiment were provided orally to the participants in their native language of Japanese. All the participants were administered the reading section of the Michigan Placement Test: Form C in order to measure their English ability. They were randomly assigned a packet of material including a participation agreement sheet and a passage to read with one of the three tasks (outlining, answering embedded questions, or reading-only). The participants read the passage at their own pace while doing the assigned tasks and produced written recalls of the passages without looking back at the passage immediately after reading and one week later. They were also informed that they would later be tested for their comprehension. However, they were not informed of the written recall task so that they would read naturally and not attempt to memorize the passage. Table 1 summarizes the sample size of each condition for both the immediate and delayed recall.

Table 1
*Grouping of Participants by Task, Condition, and Intervals*

| | Think-aloud | | | Non-think-aloud | | |
|---|---|---|---|---|---|---|
| | Control | Embedded questions | Outlining | Control | Embedded questions | Outlining |
| Immediate Recall | 12 | 10 | 9 | 10 | 12 | 11 |
| Delayed Recall | 10 | 11 | 9 | 10 | 10 | 9 |

The participants in the think-aloud group verbally reported their cognitive processes to the language laboratory while reading and completing the task, whereas those in the non-think-aloud group remained their classrooms. They received the think-aloud instruction and then conducted a short practice section. Participants were first shown a brief demonstration of the technique and then they were given a practice passage. The practice passage was different from the test passage and much shorter. The participants produced verbal reports about their thinking generally, in their first language, while reading a passage so that their report would not be constrained by their limited production skills. They were instructed to verbalize whatever they were thinking as their thoughts naturally came to mind while reading and doing the assigned tasks, such as comments on language, content, task or whatever they associate with them. As they were also asked explain how they make decision when they provide written forms of responses to the while-reading tasks, it is possible to refer to the think-aloud method as what Bowles and Leow (2005) call "metalinguistic think-aloud". Reading and task completion times were measured for all groups.

### *3.4 Analysis*

Written recall analysis was adopted to assess reading comprehension because it permits the qualitative analysis of the effects of particular interventions on reading comprehension and it also permits the analysis of relationship between reading comprehension and reading processes. Before analyzing the recall data, the passage was propositionally analyzed according to the procedure proposed by Bovair and Kieras (1985) and divided into propositions by two trained raters. A proposition was defined in this study as a semantic unit consisting of a predicate and one or more than two arguments. Using this definition, a list of propositions was created. According to this analysis, the text was made up of 185 propositions. This list was used to score the amount of information contained in the participants' recall products. Thus, the maximum score of the total recall was 185 points.

Thus, each participant's recall protocol was analyzed in terms of the total number of recalled idea units. Twenty-five percent of the data were scored by two trained raters and inter-rater reliabilities as calculated by dividing the number of propositions both raters agreed upon by the total number of propositions in the passage were found to be .90-.95. All disagreements were resolved by discussion and another round of scoring by the two raters took place. After determining that the inter-rater reliability was sufficiently high, the rest of the data were analyzed by one rater. Group means were obtained for the total proportion of recalled propositions (the total number of recalled propositions was divided by the total number of propositions for each passage).

## 4 Results

In order to measure the effects of think-alouds and task type on passage recall (immediate and delayed), the mean proportion and standard deviations of the total ideas recalled were calculated (see Table 2). The raw scores were submitted to two separate two-way ANOVAs using between-groups factors of reading condition and task type. As totally three two-way factorial ANOVAs were conducted with the same participants in this study, a Bonferroni adjustment was made, i.e., the alpha level of .05 was divided by 3 and an adjusted alpha level of .017 was employed in order to avoid making a Type I error.

The results, which are shown in Tables 3 and 4, revealed no statistically significant difference for both passage recalls between the two reading conditions (think-aloud and non-think-aloud). No

significant differences in while-reading tasks were shown for either immediate recall or delayed recall. There was no significant interaction between reading condition and task type. These results suggest that those participants who thought aloud during a reading task recalled the passage as well as those who did not, regardless of the type of task they engaged in during reading. It was also suggested that the long-term retention of the passage was not affected by either reading condition or task type.

Table 2
*Mean Proportion (Standard Deviation) of the Recalled Ideas by Task Type and Reading Condition*

| | Immediate Recall | | | Delayed Recall | | |
|---|---|---|---|---|---|---|
| | Control | Embedded | Outline | Control | Embedded | Outline |
| Think-aloud | 40.50 | 39.10 | 33.18 | 18.30 | 20.82 | 19.40 |
| | (16.38) | (13.64) | (19.27) | (9.42) | (15.18) | (10.97) |
| Nonthink-aloud | 40.30 | 34.00 | 39.30 | 18.60 | 15.80 | 16.56 |
| | (20.53) | (15.77) | (17.78) | (13.02) | (4.83) | (16.68) |

Table 3
*Two-Way Factorial ANOVA on Immediate Recall by Task Type and Reading Condition*

| Source | *SS* | *df* | *MS* | *F* | *p* |
|---|---|---|---|---|---|
| Task (T) | 234.21 | 2 | 117.11 | .392 | .677 |
| Condition (C) | 237.32 | 1 | 237.32 | .795 | .376 |
| T × C | 108.40 | 2 | 54.20 | .182 | .834 |
| Error | 17901.75 | 58 | 298.36 | | |

Table 4
*Two-Way Factorial ANOVA on Delayed Recall by Task Type and Reading Condition*

| Source | *SS* | *df* | *MS* | *F* | *p* |
|---|---|---|---|---|---|
| Task (T) | 2.28 | 2 | 1.14 | .008 | .993 |
| Condition (C) | 95.00 | 1 | 95.00 | .630 | .431 |
| T × C | 72.93 | 2 | 36.47 | .242 | .786 |
| Error | 8144.36 | 54 | 150.82 | | |

In order to investigate the effect of think-aloud task and different types of learning tasks on the reading time (seconds), the mean proportion and standard deviations of the time the participants spent for reading and task completion were calculated. Table 5 shows the descriptive statistics.

Table 5
*Means (SDs) of Reading Time by Task Type and Reading Condition*

| | Control | Embedded Questions | Outlining |
|---|---|---|---|
| Think-aloud | 1476.00 (884.21) | 1523.20 (455.56) | 1386.00 (220.56) |
| Non-think-aloud | 931.40 (374.34) | 1014.42 (345.49) | 1084.91 (352.15) |

*Note.* Time is presented in seconds. The reading time includes the time participants spent completing the processing tasks.

Whether verbalization as well as task differences affect the mean proportion of reading time was examined with a two-way ANOVA using two between-groups factors of reading condition and task type. As shown in Table 6, a statistically significant effect of reading condition for reading time was found ($p$ = .0029). Thus, the think-aloud group took considerably more time than the control group. This suggests that verbalization had a significant effect on the amount of time it took participants to complete the reading task.

The proportion of variance accounted for in the overall ANOVA design was estimated using $eta^2$. The results indicated that the main effect for reading condition accounted for 18% for the variance of the time, which indicates its importance in the context of this study.

Table 6
*Two-Way ANOVA on Reading Time by Task Type and Reading Condition*

| Source | *SS* | *df* | *MS* | *F* | *p* | $eta^2$ |
|---|---|---|---|---|---|---|
| Task (T) | 46253.38 | 2 | 23126.69 | .09 | .913 | .003 |
| Condition (C) | 3226029.25 | 1 | 3226029.25 | 12.76 | .001 | .178 |
| T × C | 176797.85 | 2 | 88398.92 | .35 | .706 | .010 |
| Error | 14663139.83 | 58 | 252812.76 | | | .810 |

## 5 Follow-up analysis

The results above demonstrated no statistically significant effects of think-alouds during L2 reading comprehension, in terms of the amount of information recalled. However, they did not show whether there is any difference in the performance on the assigned while-reading tasks (i.e. outlines and responses to embedded questions). Therefore, the purpose of this analysis was to examine the performance of written products they created for their while-reading tasks and determine whether thinking-aloud affects the performance on while-reading tasks.

### *5.1 Method*

#### *5.1.1 Participants*

Out of the original pool of 28 participants' answers to embedded questions that they produced during reading, 13 participants' answers (7 from the think-aloud group and 6 from the non-think-aloud group) were analyzed. The responses to the embedded questions were available only from these participants because their data were collected in order to carry out the follow-up study.

For the participants who created outlines, outlines from 15 participants (8 products from the think-aloud group participants and 7 products from the non-think-aloud group participants) out of the original pool were examined for the same reason stated above. No significant differences between the two groups suggests that participants' proficiency levels in the two groups did not differ greatly ($p > 0.5$).

#### *5.1.2 Materials*

Answers to five embedded questions that participants gave during reading in the previous study were examined. The questions were answered either in English or Japanese in both reading conditions. For the outlining condition, outlines created by participants during reading were investigated. Outlines were created based on the instruction explained either in English or Japanese in both reading conditions.

#### *5.1.3 Scoring procedures*

Written answers were scored based on a template for scoring agreed between two raters. After discrepancies about possible answers between the two raters were discussed, an agreed version of the template was created. The maximum number of points was 5 points. Each participant's answer was scored holistically by two trained raters. After scoring two participants' responses, all disagreements were resolved by discussion and another round of scoring took place. The inter-rater reliabilities, as calculated by dividing the number of items both raters agreed upon by the total number of questions, were found to be .95.

Outlines were scored based on the criteria for scoring agreed between two raters (see Appendix). After discrepancies about the outline reflecting the organization of the passage and choice of expressions between the two raters were discussed, an agreed version of the criteria was created. The maximum number of points was 5 points. Each participant's outline was holistically scored by two trained raters. After scoring two participants' outlines, all disagreements were resolved by discussion and another round of scoring by the two trained raters took place. The inter-rater reliabilities, as

calculated by dividing the number of sub points both raters agreed upon by the total number of sub points were found to be .96.

### *5.2 Results*

To measure the effects of thinking-aloud on performance of while-reading task (embedded questions and outlines), the mean proportion and standard deviations of the scores of both tasks were calculated (see Table 7). Descriptive statistics indicate that the difference between the think-aloud group and non-think-aloud group is more marked with outline tasks than embedded questions. The raw scores were submitted to a two-way factorial ANOVA using two between-groups factors of reading condition and task type. The results, which are shown in Table8, indicate that a significant effect was found for task types. It is suggested that participants' scores on embedded questions are higher than those on outlines. Although the reading condition did not reach significance, there is a statistical tendency that suggests a meaningful relationship between the verbalization factor and performance of while-reading task.

The proportion of variance accounted for in the overall ANOVA design was estimated using $eta^2$. The results indicated that the main effect for task accounted for 17% for the variance of the performance in while-reading tasks. This suggests the importance of the task effect.

Table 7
*Means (SDs) of Performance on Embedded Questions and Outlines by Reading Condition*

| | Embedded question | Outline |
|---|---|---|
| Think-aloud | 4.25 ( 1.07) | 2.84 ( 1.40) |
| Non think-aloud | 4.50 ( .32) | 4.00 ( .78) |

Table 8
*Two-Way ANOVA on Performance on While-reading Task by Task Type and Reading Condition*

| Source | *SS* | *df* | *MS* | *F* | *p* | $eta^2$ |
|---|---|---|---|---|---|---|
| Task (T) | 6.29 | 1 | 6.29 | 6.04 | .022 | .174 |
| Condition (C) | 3.43 | 1 | 3.43 | 3.29 | .082 | .094 |
| T × C | 1.42 | 1 | 1.42 | 1.37 | .254 | .039 |
| Error | 24.99 | 24 | 1.04 | | | .671 |

## 6 Discussion

The first research question which asked whether the effects of think-alouds interact with differences in while-reading tasks on text comprehension in terms of the number of recalled ideas was not supported. This is because neither think-aloud effect nor task type effect influenced reading comprehension. The reactivity of think-alouds was not shown possibly because of lack of additional cognitive demand which could be created by thinking-aloud. Comprehending a passage written in L2 may not be cognitively demanding enough for participants' recalls to be affected by thinking-aloud, compared to other L2 learning tasks. Furthermore, thinking aloud may have possibly helped participants comprehend and remember the text by reading aloud difficult parts repeatedly. Such possible facilitative effects of think-alouds may have cancelled out the negative effect of additional verbalization task.

One of the reasons why the while-reading tasks did not influence reading comprehension may be that potential differences in task demand seem to have been cancelled out by the mismatch between the task types and the participants' linguistic proficiency level: the two tasks may not have been different from each other in terms of cognitive demand prompted by task requirement. Content analysis of think-aloud protocols in Yoshida (2007) showed although outlining and answering embedded questions induced different types of processes, such influence was not reflected in the amount of information recalled because of participants' limited linguistic proficiency. As different types of facilitative effects of outlining and answering embedded questions have been reported mainly in the L1 literature, such

impact may not appear in L2 reading unless readers' language proficiency is high enough to be affected by conceptual elaboration being promoted through the while-reading tasks. Below this level, interventions at a linguistic level such as textual elaboration or simplification might be more influential.

The second research question asking whether thinking-aloud facilitates or inhibits text comprehension in terms of the amount of information recalled was answered negatively. In other words, reactivity of think-aloud was not found on text comprehension in terms of the number of recalled ideas; and non-reactivity was confirmed by the lack of difference between the think-aloud group and non-think aloud group on their delayed recalls. These results also support the assumption in reading studies that think-aloud protocols can yield rich data concerning the flow of information through working memory and can do so without intruding significantly on the comprehension process itself (cf. Ericsson & Simon, 1984, 1993).

In addition to a possible reason for the lack of reactivity pointed out above, it is also possible to assume that verbalizing ones' thoughts to language laboratory may not have been demanding enough to impede text comprehension. Because the think-aloud session in this study was group-administered at the language laboratory, participants may have been more relaxed than when individually conducted and participants did not have to feel much pressure about their performance on the think-aloud task. Content analysis of the type of information in their verbal report and comparing different types of think-aloud protocols may clarify how differently participants perceive and carry out think-alouds.

The third research question asking whether thinking-aloud increase the amount of time on task was answered affirmatively. This is consistent with the previous findings and assertions made in the previous study: the rate of thinking has to be slowed down in order to allow for the additional time required for verbalization of thought (Bowles & Leow, 2005; Sanz, et al., to appear).

The follow-up analysis investigated whether thinking-aloud affect the performance produced in a written form as a while-reading task. The tendency for statistical significance for reading condition on the performance on outlines and scores of embedded questions may suggest that while-reading tasks such as outlining and answering questions can be affected by thinking-aloud. Further studies are needed to confirm or disconfirm this speculation. If it is confirmed, it is suggested that cognitively demanding while-reading tasks are more likely to suffer from reactivity of think-aloud protocols. This is consistent with Jourdenais's (2001) assertion that think-aloud acts as an additional task, especially when learners engage in cognitively complex tasks such as written production. It was argued that the think-aloud data collection method itself acts as an additional task, based on Jourdenais (1998) which showed that thinking-aloud restrained L2 learners from using the targeted linguistic forms in a writing task.

Descriptive data hinted the possibility that outlining might be affected by thinking-aloud than answering embedded questions. The inclusion of larger sample size may increase difference in cognitive demand between the two while-reading tasks under the think-aloud condition and show an interaction between the task type and verbalization condition. As the two while-reading tasks used in the present study are assumed to facilitate different type of reading processes and require different degrees of cognitive demand, a different degree of reactivity of thinking-aloud might be shown, if clear distinction between two while-reading tasks, in terms of cognitive demand was empirically proved. Further investigation into the interaction between task type and reading condition (i.e. think-aloud or not) is required, if we are to identify how differences in task demand relate to reactivity effects of think-alouds.

## 7 Conclusion

The present study was designed to address the issue of reactivity in L2 reading research, in relation with type of task which participants engage in during reading. To determine the effects of think-alouds while reading in L2 and engaging in a while-reading task, reading time, the amount of information recalled about the passage, and performance on the while-reading tasks were compared between the think-aloud group and non-think-aloud group. The results indicated think-aloud protocols did not cause reactivity effects on L2 reading in terms of recalled ideas but could possibly affect performance produced in a written form as a while-reading task.

The findings about reactivity in the present study have shed new insights into L2 reading research as well as having built on the previous reactivity studies in SLA studies. The finding that

thinking-aloud is not reactive on L2 reading but possibly reactive on complex tasks such as while-reading tasks which involve writing is meaningful. This information is important enough to consider an optimal L2 reading situation, considering the concern that L2 reading is inherently cognitively demanding due to linguistic constraints. Thus, the cognitive resources allocated to the reading task *per se* would be limited in L2 reading.

The present study also raised questions for future studies. As the purpose of using think-aloud protocols in reading research is to observe on-line processes, which usually requires substantial training and individual administration rather than group administration, future L2 reading research would benefit more from investigating differences in reactivity effects depending on the type of think-aloud (e.g. well-trained think-alouds vs. less trained think-alouds; group administration vs. individual administration). Furthermore, reactivity issues should be discussed by examining processes readers engage in rather than final products created by readers. Research into the effects of think-alouds on individual reader's processes would provide valuable information. Theses studies will help researchers better understand the conditions under which the think-aloud protocol yields the best insights into reading processes.

Consistent with the caution implicated by previous research (cf. Jourdenais 1998), this study also hinted at the possibility that the thinking-aloud might affect task performance, depending on types of task. We must be careful about claims we make regarding what it is that we are measuring. Despite its limitations, however, the think-aloud analysis is still a valuable on-line tool for studying L2 reading processes. As Yamashita (1999) suggested, whether good verbal protocols are obtained for research purposes largely depends on how the think-aloud is used in a particular study. Even though reactivity is possibly observed in some performances, the think-aloud protocols can provide valuable information, as long as it reflects common patterns of processes between the experimental conditions of the study. The use of multiple approaches for data collection can also compensate for the lack of validity and reactivity of the think-aloud protocols. The use of multiple sources can reduce the risks of making generalization based on limited information.

## Appendix

*Scoring criteria for outlines*

| Criteria | point |
|---|---|
| 1st main point: the present situation of deforestation | 1 |
| 2nd main point: causes for deforestation | 1 |
| 3rd main point: negative effects | 1 |
| Subtopics under the 2nd main point: small farmers and land owners | 1 |
| Subtopics under the 3rd main point: land sliding and global warming | 1 |
| Total | 5 |

## References


Afflerback, Peter (2000). Verbal reports and protocol analysis. In Michael L. Kamil, Peter B. Monsethal, P.David Peason, & Rebecca Barr (Eds.), *Handbook of Reading Research* (Vol. 3). Mahwah, NJ: Erbaum.

Alderson, J. Charles (1990). Testing reading comprehension skills (Part one). *Reading in a Foreign Language, 6*(2), 425-438.

Block, Ellen (1992). See how they read: Comprehension monitoring of L1 and L2 readers. *TESOL Quarterly, 26*(2), 319-343.

Baumann, James F., Seifert-Kessell, Nancy, and Jones, Leah. A. (1992). Effect of think-aloud instruction on elementary students' comprehension monitoring abilities. *Journal of Reading Behavior, 24,* 143-172.

Bovair, Susan, & Kieras, David E. (1985). A guide to prepositional analysis for research on technical prose. In Bruce K. Britton & J. B. Black (Eds.), *Understanding expository text* (pp. 315-362). Hillsdale, NJ: Lawrence Erlbaum.

Bowles, Melissa A. (2008). Task type and reactivity of verbal reports in SLA: A first look at an L2 task other than reading. To appear in *Studies in Second Language Acquisition*, 30 (4).

Bowles, Melissa A. & Leow, Ronald P. (2005). Reactivity and type of verbal report in SLA research methodology. *Studies in Second Language Acquisition, 27,* 415-440

Chi, Michelene T. H., de Leeuw, Nicholas, Chiu, Mei-hung, & La Vancher, Christian (1994). Eliciting self-explanations improves understanding. *Cognitive Science, 18,* 439-477.

Coté, Nathalie, Goldman, Susan R., & Saul, Elizabeth U. (1998). Students making sense of Informational text: relations between processing and representation. *Discourse Processes 25-1,* 1-53

Ellis Nick (2001) Investigating form-focused instruction. *Language Learning, 51,* 1-46.Fletcher, C. R. (1986). Strategies of the allocation of short-term memory during comprehension. *Journal of Memory and Language, 25,* 43-58.

Ericsson, K. Anders (2002). Towards a procedure for eliciting verbal expression of non verbal experience without reactivity: Interpreting the verbal overshadowing effect within the theoretical framework for protocol analysis. *Applied Cognitive Psychology, 16,* 981-987.

Ericsson, K. Anders & Simon, Herbert A. (1984). *Protocol analysis: Verbal reports as data*. Cambridge, MA: MIT Press.

Ericsson, K. Anders & Simon, Herbert A. (1993). *Protocol analysis: Verbal reports as data* (Rev.ed.). Cambridge, MA: MIT Press.

Fletcher, Charles R. (1986). Strategies of the allocation of short-term memory during comprehension. *Journal of Memory and Language, 25,* 43-58.

Horiba, Yukie (2000). Reader control in reading: Effects of language competence, text type, and task. *Discourse Processes, 29*(3), 223-267.

Jourdenais, Reneé (1998). The effects of textual enhancement on the acquisition of the Spanish *preterit and imperfect*. Unpublished Ph.D. dissertation, Georgetwon University, Washington, DC.

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

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

Leow, Ronald P., & Morgan-Short, Kara (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

Loxterman, Jane A., Beck, Isabel L., & McKeown, Margaret G. (1994). The effects of thinking aloud during reading on students' comprehension of more or less coherent text. *Reading Research Quarterly,* 29, 352-367.

Lyons, William (1986). *The disappearance of introspection.* Cambridge, MA: MIT Press.

Matsumoto, Kazuko (1993). Verbal-report data and introspective methods in second language research: State of the art. *RELC Journal, 24*(1), 32-60.

Mikulecky, Beatrice S., & Jeffries, Linda (1996). *More reading power*. US: Longman. *Psychology Review, 1*(2), 113-145.

Nisbett, Richard E., & Wilson, Timothy DeCamp (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84,* 231-259.

Sachs, Rebecca, & Polio, Charlene (2007). Learners' uses of two types of written feedback on an L2 writing revision task. *Studies in Second Language Acquisition, 29*(1), 67-100.

Sachs, Rebecca, & Suh, Bo-Ram (2007). Textually enhanced recasts, learner awareness, and L2 outcomes in synchronous computer-mediated interaction. In Alison. Mackey (Ed.), *Conversational interaction in second language acquisition: A collection of empirical studies* (pp. 197-227). Oxford: Oxford University Press.

Sanz, Christina, Lin, Hui-Ju, Lado, Beatriz, Bowden, Harriet W., Stafford, Catherine A. (2009). Concurrent verbalizations, pedagogical conditions, and reactivity: Two CALL studies. To appear in *Language Learning,* 59 (1).

Selinger, Herbert W (1983). The language leaner as linguist: Of metaphors and realities. *Applied Linguistics*, 4, 179-191.

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

Yoshida, Mami (2007). *The effects of task, text, and proficiency on second language reading.* Unpublished doctoral dissertation. Temple University

Zwaan, Rolf A., & Brown, Carol M. (1996). The influence of language proficiency and comprehension skill on situation-model construction. *Discourse Processes, 21*, 289-327.

# Selected Proceedings of the 2007 Second Language Research Forum

## edited by Melissa Bowles, Rebecca Foote, Silvia Perpiñán, and Rakesh Bhatt

**Cascadilla Proceedings Project Somerville, MA 2008**



ISBN 978-1-57473-425-6 library binding




### Ordering information

Orders for the library binding edition are handled by Cascadilla Press.
To place an order, go to www.lingref.com or contact:

Cascadilla Press, P.O. Box 440355, Somerville, MA 02144, USA
phone: 1-617-776-2370, fax: 1-617-776-2271, e-mail: sales@cascadilla.com

### Web access and citation information

This entire proceedings can also be viewed on the web at www.lingref.com. Each paper has a unique document # which can be added to citations to facilitate access. The document # should not replace the full citation.

This paper can be cited as:

Yoshida, Mami. 2008. Think-Aloud Protocols and Type of Reading Task: The Issue of Reactivity in L2 Reading Research. In *Selected Proceedings of the 2007 Second Language Research Forum*, ed. Melissa Bowles, Rebecca Foote, Silvia Perpiñán, and Rakesh Bhatt, 199-209. Somerville, MA: Cascadilla Proceedings Project.

or:

Yoshida, Mami. 2008. Think-Aloud Protocols and Type of Reading Task: The Issue of Reactivity in L2 Reading Research. In *Selected Proceedings of the 2007 Second Language Research Forum*, ed. Melissa Bowles, Rebecca Foote, Silvia Perpiñán, and Rakesh Bhatt, 199-209. Somerville, MA: Cascadilla Proceedings Project. www.lingref.com, document #1745.