

14-818
2
03
54772210



This is to certify that the
thesis entitled

REFORMULATION, NOTICING, AND SECOND LANGUAGE
WRITING

presented by

REBECCA RAEWYN SACHS

has been accepted towards fulfillment
of the requirements for the

Master of Arts degree in Teaching English to Speakers
_____ of Other Languages _____

A handwritten signature in dark ink, appearing to be "C. R.", written over a horizontal line.

Major Professor's Signature

A handwritten date "May 8, 2003" written in dark ink over a horizontal line.

Date

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
JUL 06 2005 MAY 08 2007	JAN 10 2005	
JAN 13 2007		

REFORMULATION, NOTICING, AND SECOND LANGUAGE WRITING

By

Rebecca Raewyn Sachs

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

2003

ABSTRACT

REFORMULATION, NOTICING, AND SECOND LANGUAGE WRITING

By

Rebecca Raewyn Sachs

Proposed methods of improving corrective feedback in L2 writing classes often suggest increasing students' involvement in noticing and analysis, assuming that search, evaluation, depth of processing, and self-sufficiency will help to promote interlanguage development. In an exploratory study of two learners who revised their essays after comparing them with reformulations, Qi and Lapkin (2001) found that quality of noticing was directly related to L2 writing improvement. This thesis seeks to confirm their findings quantitatively and to compare reformulation and explicit error correction with respect to the noticing they promote. It also investigates the effects of using think-aloud protocols, not only from the standpoint of veridicality and reactivity, but also with the idea that verbalization might enhance quality of noticing. In the first study of this thesis, a repeated measures design, 15 ESL learners participated in three writing conditions (error correction, reformulation, and think-aloud), counterbalanced to control for writing topic and order of condition. Their essays and revisions were analyzed to compare changes in accuracy (with possible evidence of noticing) among the three conditions. 54 ESL learners then participated in a similar study with a non-repeated measures design. In both studies, the students in the error correction condition consistently produced the most accurate revisions. The findings suggest avenues for further research which will give us insight into feedback processing, quality of noticing, and research methodology.

ACKNOWLEDGMENTS

This thesis would not have been possible without the help and support of a number of people. I would like to thank all of the students who participated in the studies; Jonathan DeHaan, Cathy Mazak, Andy McCullough, Suzanne Bonn, Cathy Allen, and Gigi Ignatowski for letting me perform research in their classes; and Professor Susan Gass for being on my M.A. committee and giving me a great deal of insightful advice. I would especially like to thank Professor Charlene Polio for her hours of data coding and comparisons, help with statistics and think-alouds, detailed comments on drafts, classes that prepared me for this experience, her own invaluable research experience, advice, encouragement, and ideas that shaped the study, for being fun to work with, and for not coding this sentence 8.

TABLE OF CONTENTS

LIST OF TABLES.....	vi
LIST OF FIGURES.....	viii
CHAPTER 1	
INTRODUCTION.....	1
1.1 Corrective Feedback.....	1
1.2 Summary of Qi and Lapkin (2001) and Research Questions.....	4
CHAPTER 2	
LITERATURE REVIEW.....	9
2.1 Research on Noticing.....	9
2.2 Reformulation.....	15
2.2.1 Output hypothesis.....	16
2.2.2 Negative and positive evidence.....	17
2.2.3 “Deeper feedback” than with error correction: Focus on both meaning and form.....	19
2.2.4 The element of search: Increased cognitive load.....	19
2.3 Quality of Noticing and Think-Aloud Protocols.....	21
2.3.1 Reactivity and nonveridicality.....	23
2.3.1.1 Reactivity.....	24
2.3.1.2 Nonveridicality.....	27
2.3.2 Effects of verbalizations with different kinds of tasks.....	33
2.3.3 Factors causing reactivity.....	36
2.3.4 Applicability to L2 research.....	38
2.3.5 Task characteristics.....	40
2.3.6 Training and instructions.....	41
2.3.7 Experimenter influence: Social interaction.....	45
2.3.8 Verbal protocols in an L2.....	50
2.4 Summary.....	53
2.5 Hypotheses.....	57
CHAPTER 3	
STUDY 1 (REPEATED MEASURES DESIGN).....	62
3.1 Participants (Study 1: Repeated Measures).....	62
3.2 Design (Study 1: Repeated Measures).....	63
3.3 Results (Study 1: Repeated Measures).....	69
3.4 Analysis of Think-Alouds.....	77
3.4.1 Association between noticing (and quality of noticing) and correction.....	79
3.4.2 Quality of noticing and noticing the gap.....	80
3.5 Problems Leading to Study 2 and Rationale for Modifications in Design.....	81

CHAPTER 4	
STUDY 2 (NON-REPEATED MEASURES DESIGN).....	85
4.1 Participants (Study 2: Non-Repeated Measures).....	85
4.2 Results (Study 2: Non-Repeated Measures).....	85
CHAPTER 5	
DISCUSSION.....	89
5.1 Discussion of Research Questions.....	89
5.1.1 Research question 3: Do students notice more when comparing their essays to reformulated versions as opposed to versions with explicit error corrections?.....	89
5.1.2 Research question 4: Does the use of think-aloud protocols affect the number of linguistic features that students notice and that subsequently make their way into the final version of the written text?...92	92
5.1.3 Research question 1: What do L2 learners notice as they compare their text to a reformulated version while thinking aloud?.....	96
5.1.4 Research question 2: How is noticing related to revision changes completed after comparing the original and reformulated versions of story?.....	98
5.2 Implications for Research Methodology.....	100
5.2.1 Problems with attempts to distinguish between noticing of different qualities.....	100
5.2.2 Additional effects of error type on the construct “quality of noticing”.....	103
5.3 Further Research.....	105
5.4 Implications for Pedagogy.....	106
APPENDICES.....	109
Appendix A: Counterbalance Chart for Repeated Measures Study.....	110
Appendix B: Figure 1. Writing Prompt A.....	111
Appendix C: Error Classification System.....	112
Appendix D: In-Class Instructions.....	115
Appendix E: An Example of Error Coding, An Example of Explicit Error Corrections, An Example of a Story and its Reformulation, and An Example of an Error Tally Sheet (Student A)	116
Appendix F: Think-Aloud Instructions.....	121
Appendix G: An Example of Columns Format.....	122
Appendix H: Guidelines for Division into T-units.....	124
Appendix I: Coding System for Changes in Accuracy.....	126
Appendix J: 3-Tiered Coding System for the Quality of Noticing Related to Each Error, Based on Think-Aloud Data.....	130
Appendix K: Selected Quotations from the Post-Study Debriefings.....	132
REFERENCES.....	134

LIST OF TABLES

Table 1: Three-day sequences of the three experimental conditions.....	64
Table 2: Comparison of conditions with regard to evidence of noticing (in percentage form).....	70
Table 3: Comparison of times with regard to evidence of noticing (in percentage form).....	71
Table 4: Comparison of times with regard to evidence of noticing: Friedman Test of ranked percentages.....	71
Table 5: Comparison of conditions with regard to evidence of noticing: Friedman Test of ranked percentages.....	72
Table 6: Comparison of the Error Correction and Reformulation conditions: Wilcoxon Signed Ranks Test.....	73
Table 7: Comparison of the Think-Aloud and Reformulation conditions: Wilcoxon Signed Ranks Test.....	74
Table 8: Comparison of the Think-Aloud and Error Correction conditions: Wilcoxon Signed Ranks Test.....	74
Table 9: Comparison of conditions with regard to complete correction (in percentage form).....	75
Table 10: Comparison of conditions with regard to complete correction: Friedman Test of ranked percentages.....	76
Table 11: Comparison of times with regard to complete correction (in percentage form).....	76
Table 12: Associations in the think-aloud data between noticing and correction and between “high quality” noticing and correction.....	81
Table 13: Comparison of conditions with regard to evidence of noticing (in percentage form).....	87
Table 14: Kruskal-Wallis nonparametric test.....	87
Table 15: Percentages of correction for individual error types compared across condition (in percentage form, problematic).....	104

Table 16: Counterbalance Chart for Repeated Measures Study.....	110
Table 17: An Example of an Error Tally Sheet (Student A).....	120

LIST OF FIGURES

Figure 1: Writing Prompt A.....111

Chapter 1

INTRODUCTION

1.1 Corrective Feedback

In responding to L2 writing, teachers often use corrective feedback as one way of helping their students to focus on form and notice their linguistic problems. For their part, learners may believe that error correction helps them to identify and resolve grammatical difficulties and write more like native speakers, and ESL university students in particular may see error-free writing as crucial to their academic success and therefore value and expect error correction. Theoretically, as one specific form of consciousness-raising, it seems as though corrective feedback should be helpful. According to Long (1998) and others, positive evidence in the form of meaningful input may not be enough for successful second language acquisition (SLA). L2 learners also need negative evidence in order to show them what is *not* possible in a language and to limit their overgeneralizations. Furthermore, the output hypothesis suggests that students might be especially inclined to notice teachers' feedback when it is related to language that they have already attempted to produce (Swain, 1995). Although Truscott (1998) has argued that conscious awareness is not necessary for learning and that noticing is not helpful for developing L2 competence, he does mention that noticing might help learners develop metalinguistic knowledge. In turn, metalinguistic knowledge and the noticing of teachers' feedback might be particularly helpful in writing since writing provides

opportunities for students to step back and analyze the language they have put down on paper, an argument also put forth by Qi and Lapkin (2001).

Practical problems involved with corrective feedback, however, have led many to question its effectiveness. Truscott (1996) rightly points out that grammar acquisition is not a sudden discovery and that the memorization of explicit rules may be superficial and transient. Some also argue that, while content- and organization-based feedback can be helpful in terms of developing students' writing ability, grammar correction does not seem to promote learners' interlanguage development. Others note that corrective feedback can have negative effects on students' affect and their ability to revise their papers comprehensively and meaningfully. Truscott (1996) goes so far as to assert that grammar correction is not only ineffective, but even counterproductive, and that it should therefore be abandoned. He points out that students must be able to understand their teachers' explanations, know what to focus on, be motivated, think about their errors in future writing, and not feel so overwhelmed that they become less willing to take risks with complex structures. The situation is not helped by the fact that, according to Zamel (1985), teachers often respond inconsistently, vaguely, and even somewhat arbitrarily to students' texts. Accordingly, Qi and Lapkin (2001) argue that teachers' feedback (in the form of written error correction) "does not provide optimal conditions to help learners *notice* their errors, i.e., the gap between their IL and TL when they receive and process the feedback" (p. 280).

With concerns like these in mind, some researchers have suggested possible ways to make the provision and utilization of corrective feedback a more viable and worthwhile endeavor. Corder (1981), for instance, proposed that teachers should modify

their feedback so that students can approach it as a problem-solving activity. Makino (1993) likewise asserted that the provision of cues instead of explicit corrections could make students more active participants in the process. Behind both of these suggestions is the idea that cognitive activity and the development of self-sufficiency are important. In addition, it might not be unreasonable to think that the element of active search and its relation to depth of processing may also be factors, as some studies of vocabulary acquisition have indicated. In Laufer and Hulstijn's (2001) explanation of task-induced involvement, they noted that a "higher involvement load" in relation to need, search, and evaluation can have a positive effect on learning. For example, in a study by Hulstijn, Hollander, and Greidanus (1996), when learners looked up words in a dictionary, they had higher retention for the new vocabulary items than learners who were simply provided with the words' meanings in marginal glosses.

Both of the latter two articles focused on vocabulary acquisition, but perhaps the ideas of search and depth of processing can be extended beyond the lexicon and applied to the interpretation of corrective feedback as well. Of course, searching one's IL for an understanding of grammatical differences is not the same as searching for a word in a dictionary, which at least can be expected to provide a relatively complete, unambiguous answer. However, it can be hypothesized that if learners must actively engage their IL systems and evaluate their existing knowledge while they process their teachers' feedback, this might lead to greater uptake than the simple noting and copying of explicit error corrections. As far as SLA research is concerned, it would be helpful to find out systematic information about what L2 learners notice and how that compares to what they are able to incorporate into their own language production (Schmidt, 1990).

1.2 Summary of Qi and Lapkin (2001) and Research Questions

Qi and Lapkin (2001) approached these issues within the context of *reformulation*, which has been defined by Thornbury (1997) as a native speaker's reworking of an L2 learner's written composition in order to make the language seem as native-like as possible while keeping the content of the original intact. In a pilot study with two Mandarin-speaking learners of English, Qi and Lapkin used a three-stage writing task to investigate the relationships between noticing and a variety of other factors, including composing, processing of feedback, L2 proficiency, and L2 writing improvement on a revised text. They asked the participants to think aloud throughout the process and recorded their verbalizations on audiocassette and videotape. In Stage 1, each participant was given 30 minutes to write a story based on a picture. Four days later, after the researcher had reformulated the stories' language in order to make it sound more native-like, the participants (in Stage 2) compared their original versions to what the researcher had written and engaged in retrospective interviews to clarify what they had noticed. During these interviews, the researcher showed the participants the videotapes of the text comparison process, pausing periodically and asking the participants to explain specifically what they had been noticing at the time. In Stage 3, the participants were given the chance to revise their original versions.

Qi and Lapkin asked three research questions in particular:

- 1.) What aspects of language do L2 learners notice in/during an output-only writing condition (Stage 1 of a three-stage writing task)?

- 2.) What do L2 learners notice as they compare their text to a reformulated version of it while thinking aloud (Stage 2 of a three-stage writing task)?
- 3.) How is such noticing related to changes in the written product from Stage 1 to Stage 3 (posttest) of the L2 writing task [i.e., changes made to the revision after comparing the original and reformulated versions]?

Finding that the higher proficiency learner in their study both resolved more problems in his writing and gave reasons for accepting the reformulations at a higher rate during the think-aloud (72%, compared to 23% for the lower proficiency learner), Qi and Lapkin came to the conclusion that learners with different L2 proficiency levels differ in their ability to achieve high quality noticing, which is directly related to improvement on revisions. According to Qi and Lapkin, learners with higher L2 proficiency may not only be able to notice more about the linguistic features of their own output as they compose, but they may also be better equipped to notice the gap between their writing and a reformulated version of it. They hypothesized that this may be due to the fact that higher proficiency learners (at least judging by their higher proficiency participant) tend to accept more reformulations and also more readily verbalize the reasons behind the differences they have noticed.

Of course, the generalizability of their findings is limited by the fact that there were only two participants involved in their study, and the participants certainly differed from each other in aspects other than proficiency level. However, Qi and Lapkin did

point out that their findings appear to be in line with previous research conducted by Cohen (1983) and Swain and Lapkin (2000), showing, respectively, that intermediate and advanced learners may benefit more from reformulation than beginners do and that low proficiency L2 learners may not be able to identify errors due to limitations in their L2 knowledge. In any case, Qi and Lapkin suggested that it may be pedagogically valuable, regardless of proficiency level, to promote not just noticing, but high quality noticing. Variability in quality of noticing seems to be related to variability in the ability to revise.

These sorts of hypotheses concerning relationships between noticing, L2 proficiency, and L2 writing merit further consideration for both theoretical and practical purposes, and it seems particularly important to try to isolate the relationship between noticing (and quality of noticing) and L2 writing without including proficiency as a factor. This may especially be the case if think-aloud protocols are used as a research method to tap into the writing process. Qi and Lapkin utilized think-alouds to find out what the participants were noticing during all three stages: composing, comparing, and revising. Although verbal protocols are not necessarily inherently flawed, there are many concerns to keep in mind while employing them, and it seems clear that they may affect high and low proficiency learners to different extents. First of all, a higher proficiency learner might have the advantage of being able to verbalize his/her thoughts more easily and become less distracted while doing so. Even if both a higher and a lower proficiency learner noticed a simple verb tense error, for example, it might be more difficult for the lower proficiency learner to explain it. This difficulty associated with verbalization could divide his/her cognitive resources, possibly making the error less likely to be remembered come revision time and putting the lower proficiency learner at a disadvantage just

because of limited fluency. On the other hand, if the simple act of verbalizing something makes it more likely to be remembered, a learner with greater speaking fluency would be at an advantage. The production of a think-aloud protocol may have the potential to enhance or hinder noticing during an L2 writing task.

Besides the L2 proficiency of the participants, it would seem to depend on many other factors whether verbalization might divide cognitive resources and disrupt noticing, or whether it might draw attention to linguistic items and help learners to remember them. It is also important to realize that while noticing can be operationalized as “availability for verbal report” (Schmidt, 1990), it is possible for L2 learners to notice and understand without verbalizing, and it is possible for factors other than verbalization to influence the process. Even if it is true that what learners verbalize in a think-aloud normally corresponds to what they have noticed, verbalization is not exactly equivalent to noticing, and a cause-effect relationship cannot be claimed.

This makes it even more interesting to compare how noticing may be promoted in different writing conditions and how learners of roughly the same proficiency level are able to make improvements in their writing in each condition. Qi and Lapkin found a relationship between quality of noticing and L2 writing improvement which was also related to proficiency level. It would be worthy of note if, apart from proficiency level, we could show more support for an association between quality of noticing and revision improvements. With these questions in mind, it will be important in this thesis to review research on issues related to noticing and reformulation, such as the output hypothesis, the importance of negative and positive evidence, the value of making cognitive comparisons, focus on meaning and form, and cognitive load. Additionally, since verbal

protocols will be used, the reactivity and veridicality of that method will also be discussed. The research questions for this thesis are as follows:

RQ1: What do L2 learners notice as they compare their text to a reformulated version while thinking aloud? (corresponding to Qi and Lapkin's second research question)

RQ2: How is such noticing related to changes in the written text completed after comparing the original and reformulated versions? (corresponding to Qi and Lapkin's third research question)

RQ3: Do students notice more when comparing their essays to reformulated versions as opposed to versions with explicit error corrections?

RQ4: Does the use of think-aloud protocols affect the number of linguistic features that students notice and that subsequently make their way into the final version of the written text?

Chapter 2

LITERATURE REVIEW

2.1 Research on Noticing

Uptake is an important concern in SLA. The fact of the matter is that students cannot always be expected to transfer input to output; they might need feedback or relevant input directed explicitly towards structures for them to be able to notice and integrate new forms. Ellis (1995) maintained that even while emphasizing input and interaction in communicative language teaching, it is crucial to realize that learners might need some sort of direct intervention. Research in immersion settings in particular has indicated that learners may not succeed in acquiring certain forms, even after years of hearing them in meaningful input (Doughty & Williams, 1998; Ellis, 2001). Schmidt (1990) pointed out that it is possible for unconscious or implicit learning to happen incidentally during meaningful interaction in an immersion setting, but that adults especially might require tasks that force them to notice certain kinds of information. He speculated that one possible drawback of adults' ability to allocate attentional resources strategically is that they might not automatically be as open to other stimuli in the environment. Therefore, when they do not deliberately pay attention to redundant grammatical structures, they might not acquire them.

This phenomenon is similar to what Schmidt himself experienced as he was trying to learn Portuguese. In his case, the simple fact that certain linguistic forms were available in the input was not enough for him to be able to incorporate them into his own

language production. However, when he compared what he had reported to have noticed with what he was able to produce, the two corresponded. His conclusion was that, even if this does not prove a causal mechanism, noticing does seem closely connected to emergence in production. In his words:

Subliminal language learning is impossible.... Noticing is [necessary] for converting input to intake, [and] incidental learning [i.e. learning without consciously paying attention]... is possible and effective when the demands of a task focus attention on what is to be learned. [However], paying attention is probably facilitative, and may be necessary if adult learners are to acquire redundant grammatical features (p. 129).

It makes sense that if a second language learner (or a native speaker, for that matter) is paying attention to the message being conveyed in a language, he or she might not be aware of the form being used to convey it. Furthermore, especially given that some grammatical forms may be infrequent, non-salient, and unnecessary for understanding a message, there may be grammatical aspects of input that are not readily available to function as intake (Schmidt, 1990). Schmidt has thus proposed that conscious processing of form is necessary, and that it is a portion of what a learner has noticed that becomes intake, whether the noticing was intentional or not.

It has also been suggested that certain kinds of noticing in particular may be necessary for SLA (Schmidt & Frota, 1986). In this view, not only must learners pay attention to linguistic features of input in order for it to become intake, but they must also

notice the gap between their interlanguage (IL) output and the target language (TL) input. Klein (1986) uses the term “matching” to refer to the checking of output against an external measure, while Ellis (1995) calls it “cognitive comparison” in order to highlight the fact that learners must notice both similarities and differences between IL and TL. Importantly, Ellis notes that the process of comparing what one has noticed in input to what one is currently able to produce in output can help learners both to confirm and to disconfirm hypotheses that exist in their implicit knowledge. Other researchers have discussed related phenomena and strategies that can enhance SLA. For example, O’Malley and Chamot (1990) use the terms “selective attention” and “self-evaluation” to mean, respectively, paying attention to particular linguistic items in input while carrying out a task, and making sure that output is in accordance with internal accuracy measures. All of these strategies can help learners to restructure their interlanguage systems.

Practically speaking, the restructuring of IL due to conscious experiences such as those mentioned above seems particularly important when we remind ourselves that the ways in which learners’ IL systems are affected can also determine how subsequent linguistic data are interpreted. According to Schmidt (1990), drawing on Baars’s (1988) theory of consciousness, it is essential to keep in mind that the nervous system changes as a result of conscious experiences. New material is interpreted within an unconscious context, and it then becomes integrated into that unconscious context. This idea is helpful as far as interlanguage development is concerned since it reminds us that learning is not just about moving information into long-term memory storage. Rather, knowledge becomes part of a modified context that affects how future information is perceived and integrated.

Some have suggested that explicit knowledge might be able to facilitate this process and exert influence on implicit knowledge by means of noticing. Laufer and Hulstijn's assertion that "preparatory attention and voluntary orienting vastly improve encoding" (Laufer & Hulstijn, 2001, p. 4) seems to fit well with this. Whereas implicit knowledge is intuitive, unanalyzed, and naturally occurring, explicit knowledge is conscious, analyzed, and reportable, and it shows up in problem solving and monitoring contexts (Ellis, 1995). Ellis has pointed out that explicit knowledge usually does not turn directly into implicit knowledge because of learnability constraints; that is, when learners' IL development is not sufficiently advanced, they may not be able to integrate certain kinds of new information. However, the possession of explicit knowledge might help learners to notice forms, and if this is the case, then it is important for them to notice forms, think about what they mean, and compare those form-function mappings with their own IL systems (Ellis, 1995).

In an experiment using "consciousness-raising" tasks, Fotos (1993) also supported the idea that encouraging noticing as a cognitive strategy can help learners to develop implicit knowledge from explicit. In her study, she looked at how much learners were able to notice following consciousness-raising tasks involving interactive problem solving, and she compared this to the amount of noticing that occurred following more traditional, teacher-fronted grammar lessons. She also compared both of these groups to a control group that had not developed explicit knowledge of the forms, finding that both experimental groups performed better than the control group. Students who had gone through consciousness-raising made significant improvements in proficiency and showed themselves still to be aware of the forms in meaningful input two weeks later. Fotos

reasoned that continued awareness of forms might be a prerequisite to acquisition.

Additionally, she suggested that formal instruction might lead indirectly to acquisition after learners have made cognitive comparisons and tested their new hypotheses with regard to input and output. If we recall Schmidt's experience learning Portuguese, there are clear similarities. After he had noticed forms in subsequent communicative input, he started to produce them and develop implicit knowledge from his explicit knowledge.

In reviewing the research that has been done on form-focused instruction (FFI), Ellis (2001) noted that many studies have compared groups of learners receiving FFI with groups learning more naturalistically in order to evaluate their ultimate levels of achievement and learning rates. In general, FFI has been found to be associated with higher learning rates and ultimate achievement, and most studies seem to agree that if L2 learners are developmentally ready, they do learn the forms that they have been taught explicitly. Norris and Ortega (2000) carried out a quantitative meta-analysis of experimental studies comparing explicit and implicit instructional approaches and similarly concluded that the explicit ones tended to be more effective. It has also been suggested in SLA research that form-focused instruction might promote acquisition by providing L2 learners with expectations that can facilitate the noticing of forms in input (Ellis, 2001). Noticing does not guarantee that input will become intake, and its usefulness may depend on a learner's developmental readiness; however, if noticing truly is a prerequisite to acquisition as Schmidt maintains, then instruction that promotes noticing will presumably make acquisition more likely.

Even while noting the general agreement on this topic, it is important to mention that Schmidt's assertions are not uncontroversial; Truscott (1998), for instance, stated

that noticing may lead to the acquisition of metalinguistic knowledge, but that it does not necessarily affect the authentic, normal, spontaneous use of language (competence).

Truscott has also argued that a major problem with research whose conclusions assert the helpfulness of FFI is that tests have tapped primarily into metalinguistic knowledge.

Another caveat from Norris and Ortega (2000) is related to the fact that studies comparing explicit and implicit instruction have produced a variety of results without often having been replicated. Furthermore, it should be stressed that even though rates of learning and levels of achievement may be influenced by the type of instruction, the order of the stages of acquisition does not generally seem to change (Ellis, 2001).

This last point highlights the importance of recognizing that regardless of precisely how FFI may or may not be effective, numerous variables affect the success of any kind of instruction or feedback (Ellis, 2001). Whether or not learners notice forms and obtain intake depends on a variety of factors. As we have already seen, it is plausible that a learner's developmental stage and learnability constraints have an effect. Also significant are the materials used for instruction; the task demands; the learning environment; the frequency, perceptual salience, and complexity of the form(s) being taught; and a learner's skill level, memory, and attentional capacities, to name a few (Ellis, 2001; Schmidt, 1990; Robinson, 1995). VanPatten (1987) would also include the degree of automaticity, or a learner's ability to pay attention to both form and meaning, since the availability of cognitive resources for any given task can affect how well it is done and what parts of it sink in. All of the above points will be significant later, when the results of the particular experiments done for this thesis are discussed.

2.2 Reformulation

Concepts and issues related to noticing can also be applied to a comparison of reformulation and explicit error correction. According to Thornbury (1997), written reformulation as a technique includes “explicit form-focused, noticing-type procedures,” but the basic idea behind it is that a teacher does not simply focus on the surface features of a student’s writing (p. 328). Instead, the teacher tries to understand the student’s ideas and intentions precisely and then reformulates them, making the language seem as much as possible like that of a native speaker while keeping the content the same. Afterwards, the student can compare his or her original version with the native speaker’s version. The origins of reformulation lie in ideas proposed by Levenston (1978) and developed by Cohen (1981), two researchers who recognized its potential value; however, it also appears to be well-supported by a number of more recent theories regarding the promotion and importance of noticing, cognitive comparison, output, negative and positive evidence, depth of processing, and a focus on meaning and form.

To summarize before going into more depth, it appears that reformulation provides learners with opportunities for noticing linguistic items in input, making form-function mappings, and comparing what they have noticed with what they are currently able to produce (which, of course, is conveniently presented for them in the form of their own writing). Presumably, this input evokes personal responses since it is focused and directly related to their output. When they evaluate it with regard to their intended meanings and knowledge of rules, they may increase their awareness about their own common mistakes and, depending on readiness, get a sense for how they could have used

certain structures to express their ideas. Reformulation thus seems to be in accordance with the output hypothesis (Swain, 1985) and ideas about the importance of negative and positive evidence and of focusing on both meaning and form (Long, 1998). Since it may induce both error analysis and cognitive comparison and may require active search, one can hypothesize that it might lead to a more analytical orientation, more metalinguistic awareness, and a greater development of cognitive strategies for noticing than occurs with explicit error correction.

2.2.1 Output hypothesis

The output hypothesis suggests that the struggles learners go through while attempting to produce language output may subsequently induce them to notice particular linguistic items in input, and that this noticing might then influence what becomes intake (Swain, 1985). When learners want to convey a message, the act of language production and the occurrence of any difficulties might serve as stimuli, prompting learners to become consciously aware of their language problems and possibly pay attention to and analyze later input (Swain & Lapkin, 1995; Qi & Lapkin, 2001). This attention trigger can be activated solely based on “learner-generated input” or “autoinput,” as Fotos (1993) refers to it (p. 399), as learners go through the normal process of coming up with ways to express their ideas successfully; however, it can also take place as the result of an interlocuter’s (or reader’s) reaction to a learner’s output. Receiving native-speaker input that is related to learner output might lead to enhanced noticing of forms or even linguistic revelations and subsequent in-depth analysis. Thornbury (1997) points out that

in contrast to traditional “accuracy-to-fluency” models of instruction, reformulation is consistent with the opposite order: from fluency to accuracy (p. 328).

Furthermore, since reformulation functions as a sort of written recast – just like any other reaction to IL, but in analyzable, concrete written form – it might be particularly effective. Learners may be predisposed to notice linguistic items that they have had trouble producing or that correspond to meanings that they have not been able to produce (Johnson, 1988). This being the case, it might be helpful for teachers to show learners how to express their ideas and refine their language use *after* the learners have already made their own attempts to do so. Furthermore, it seems plausible that if teachers are able to use student-produced content and tailor their feedback to individual students’ needs and interests, learners might be more receptive (from an affective standpoint) to exploring new ways of expressing their ideas and incorporating obviously relevant linguistic forms into their writing (Frodesen, 2001).

2.2.2 Negative and positive evidence

Focused reformulation might be able to serve both as positive evidence (in a written equivalent of recasting) and as negative evidence (if learners correctly interpret it as showing them what is not allowed). To repeat, noticing is believed to be important not only for drawing attention to gaps between IL and TL and disconfirming hypotheses in students’ implicit knowledge, but also for *confirming* that IL and TL match and demonstrating positive evidence of linguistic items that have not yet been (fully) acquired (Ellis, 1995). Gass (1983) suggests that “theoretically, one could hypothesize that all sentences written by a learner would be judged grammatical by that learner since students

would not intentionally write ungrammatical sentences” (p. 279). One could argue that this is not necessarily true since students might sense that a sentence is ungrammatical but simply be unable to fix it. Besides, learners often make oversights and mistakes that they are capable of redressing themselves. However, the idea is important because at some point, an ungrammatical sentence must either sound fine to an L2 learner, or else the learner must simply leave it that way out of a lack of knowledge of how to correct it. The identification of negative evidence in relevant feedback can be particularly valuable in curbing the overgeneralization of rules (Long, 1998), and the development of learners’ interlanguage systems and intuitions may also be affected by comprehensible, contextualized positive input of increased subtlety, sophistication, and naturalness. Reformulation can provide both of these.

Johnson (1988) surmises that feedback might be most valuable when students are able to gather positive and negative evidence themselves and notice aspects of language that are appropriate for their current stages of IL development. He considers it one of the benefits of reformulation that learners may be able to notice specifically what is relevant to them. Trying to process and produce more complex language leads to IL development, and reformulation can help to provide “both the data and the incentive” for learners to make comparisons between IL and TL (Thornbury, 1997, p. 327). As learners discover particular areas in which they are lacking competence, they may become increasingly able to identify relevant negative evidence (Thornbury, 1997).

2.2.3 “Deeper feedback” than with error correction: Focus on both meaning and form

Since reformulation does not involve merely a superficial and often somewhat mechanical correction of surface errors, another aspect of its effectiveness might lie in its ability to compel students to focus on both the meanings and forms of grammatical structures (Qi & Lapkin, 2001). The use of reformulation might help teachers to encourage awareness and form-function mapping among students as a personal resource. Ellis (1995) uses the term “grammar comprehension” to mean paying attention to grammatical forms and understanding what those forms mean. This is different from “meaning comprehension,” for which a learner does not necessarily have to pay attention to redundant grammatical structures, and it is also different from simply making note of explicit grammar corrections. It makes sense that whereas pure production approaches might deal only with explicit knowledge, approaches that focus on meaning and form might help students not only to understand input, but also to obtain intake that they can integrate into their developing IL systems as they acquire target structures along with their meanings (Ellis, 1995).

2.2.4 The element of search: Increased cognitive load

With explicit error correction, students do not have to search for or notice mismatches on their own; “answers” are provided for them directly, and they simply have to make note of them. To make students more actively involved in evaluating their errors, teachers mention that they sometimes underline just the locations of mistakes and have the learners themselves attempt to figure them out. In that case, it is assumed that the students must search their minds for relevant grammatical rules or intuitions, and

even if they are not successful, the process might require more mental involvement than explicit error correction does. Robb, Ross, and Shortreed (1986) carried out a study comparing four different methods of providing feedback, ranging from relatively salient and direct to relatively non-salient and indirect. In that order, the four methods included complete correction, coded feedback, uncoded feedback with the locations of errors indicated, and marginal feedback with the errors tallied for each line of writing. Finding only negligible differences between the methods, they concluded that direct and explicit error correction did not seem to be worth the time and effort that teachers put into it and suggested that teachers should focus more on other aspects when responding to students' writing. The fact that they did not find significant differences may call into question teachers' assumptions regarding students' level of involvement in processing feedback. However, since reformulations incorporate complete correction along with additional features, it may still be instructive to investigate the kinds of search and involvement that they promote. One could perhaps argue that reformulation takes the element of search in a different direction. Learners do not have to come up with correct forms on their own as they did in the indirect and non-salient methods studied in Robb, Ross, and Shortreed; rather, they have to search for (sometimes subtle) differences and analyze how and why two versions are unlike.

The noticing, error analysis, and cognitive comparison involved in reviewing a reformulation might involve an increased cognitive load. It is important to note that *cognitive comparison* involves noticing a target linguistic item and then comparing it to IL, while *error analysis* involves noticing an IL problem in output before comparing it with a TL version (Qi & Lapkin, 2001). Reformulation can involve both processes and

may therefore help learners to monitor and be aware of what they have produced, as well as to incorporate intake and restructure their IL systems. The encouragement of noticing through reformulation might lead to more metalinguistic awareness, explicit knowledge, and an analytical orientation, and it might also speak to students' feelings for language and help them to develop their implicit knowledge and intuitions. It is not clear that explicit error correction or coding alone can do the same.

2.3 Quality of Noticing and Think-Aloud Protocols

Especially if students have not already worked on developing appropriate cognitive strategies for noticing as effectively as possible, the quality of their noticing may be variable. As Qi and Lapkin (2001) put it, noticing can be either “perfunctory” (noticing only) or “substantive” (noticing and providing reasons for differences) (p. 291). Presumably, the more in-depth and elaborate processing is done, the greater effect this will have on students' ability to revise accurately at a later time. In fact, Qi and Lapkin found in their study that when a participant verbalized a reason for an error he or she had noticed, that noticing was more likely to result in a change in the revision. From this, they asserted that “noticing *without* understanding or noticing for no articulated reason does not have the same impact on learning in L2 performance as does noticing *with* understanding” (p. 294).

While it is not possible to claim a cause and effect relationship with their data, the idea seemed to merit further investigation and provided one of the reasons behind the inclusion of think-alouds in the present study. It is possible that the necessity of

continuous verbalization can push students to rationalize in ways that they would not otherwise. The idea of reactivity is often framed in terms of negative interference with cognitive processes, and clearly, if concurrent verbalization divides the cognitive resources needed for a task, then that can be detrimental to task completion. However, it can also be hypothesized that if verbalization encourages reasoning and a greater depth of processing and leads to increased attention, it could serve as a sort of “positive interference,” actually enhancing the noticing that takes place. As a matter of fact, Ericsson and Simon (1993) noted that certain kinds of protocol instructions might improve performance in some cases and therefore have important implications for improving learning. To support this, they cited Chi et al’s (1989) finding that when subjects were studying a physics text, a greater rate of self-explanation was associated with more learning. It was also found that better students may naturally tend to use the strategy of explaining concepts out loud.

In our study, we were not so much concerned with investigating the thought processes normally involved in comparing two texts; rather, we intended to explore how rationalization and coherent description might improve quality of noticing. We wanted to encourage the sort of additional thinking that might cause learners to provide reasons for the differences they noticed. Thus, the “think-alouds” used experimentally in the main study under discussion should not be confused with “pure concurrent verbalization” as a research method. Ericsson and Simon (1993) suggest that think-alouds (as a research method) do not interfere with cognitive processes as long as participants simply report on the contents of short-term memory. In our study, however, the learners may have had to go beyond short-term memory, and other factors such as the use of an L2, experimenter

influence, and the nature of the task and task instructions might also have caused various kinds of reactivity and nonveridicality. Therefore, it will be important to review some of the issues surrounding the reactivity and nonveridicality of think-alouds and understand the effects they might have on students' noticing and subsequent abilities to revise.

2.3.1 Reactivity and nonveridicality

SLA researchers would certainly benefit from the ability to examine the processes that occur in learners' minds directly. Just because two people get the same answer to a problem does not mean that their approaches are similar or identifiable. By analogy, outside observers who have access only to learners' language output probably cannot guess which strategies they use and how frequently they use them (Cohen, 1987). There is the assumption, however, that people have access to their own internal thought processes and that they can observe and talk about perceptions, things, and ideas of which they are conscious (Gass & Mackey, 2000). Accordingly, there seems to be a growing consensus among SLA researchers that learners' own statements about how they are organizing and processing information as they carry out language tasks can be consulted as an alternative or supplement to other kinds of observations and inferences. These statements can serve as direct evidence of processes that are otherwise invisible, revealing information about the struggles students go through, the strategies they employ, the considerations that lead to decisions, the order in which they perform parts of tasks, and how individuals may be similar or different in their approaches (Hayes & Flower, 1983; Faerch & Kasper, 1987; Gass & Mackey, 2000). According to Smagorinsky

(1989), as long as researchers keep certain principles in mind, access to and analysis of learners' verbalizations can be a useful research tool.

Of course, all methods face the risk of at least some invalidity, and it has been noted that the use of think-aloud protocols in L2 research has “raced rather far ahead of the users' understanding of [its] nature and impacts” (Stratman & Hamp-Lyons, 1994, p. 89). While conceding that think-aloud protocols grant access to processing insights that may be impossible to reach by other methods, Russo, Johnson, and Stephens (1989) have also issued a challenge: to find out why and how think-alouds may be invalid, and then to improve their validity. Researchers must ask, first of all, whether participants' verbalizations of their thoughts either positively or negatively *affect* the cognitive processes they would normally use to perform a task. Then they must ask whether the protocols accurately *reflect* those cognitive processes. These two major methodological questions can be referred to as *reactivity* and *nonveridicality*, respectively.

2.3.1.1 Reactivity

Currently, most researchers consult Ericsson and Simon's 1993 book *Protocol Analysis: Verbal Reports as Data* when attempting to design studies implementing verbal protocols because it discusses at length the important question of what can be verbalized accurately without affecting underlying processes. Ericsson and Simon conceptualize human cognition as information processing and hypothesize that cognitive processes are comprised of internal states transformed in sequence. They also declare that humans have different kinds of memory storage, to the effect that information that has recently been attended to is kept in short-term memory (STM), which has limited capacity and

immediate access, but can be transferred to long-term memory (LTM), which has a large capacity, relatively permanent storage capacity, and relatively slow access time. Whereas information in STM can be accessed directly for processing and verbalizing, it is necessary to transfer information from LTM to STM before verbalizing about it (Ericsson & Simon, 1987).

According to Ericsson and Simon (1987, 1993), the way in which a task forces information to be processed affects what enters memory storage and what is reported. Three types of verbalization can be identified. At the first level, there are no intermediate processes; participants simply report on processes that are already orally encoded. For example, in solving an anagram (unscrambling letters to make a word), participants can simply verbalize the different combinations they are trying in their heads. Since the rate of silent speech has been found to be similar to the rate of overt vocalization (simply talking aloud), this can probably be done without requiring more time. At the second level, there are intermediate processes that involve putting information into an oral code for the purpose of reporting, but no new information is required. For instance, when doing a Raven's matrix, a participant must find a visual pattern in a 3x3 array of cells with one section missing and then complete the pattern by selecting from a group of alternatives. Concurrent verbalization would presumably cause this to take more time since information is maintained in attention while it is verbalized and subsequent states begin only when verbalization is complete. Despite this time difference, though, there should be no interference with underlying processes at the first two levels. Level three, on the other hand, involves more than simple recoding since information in STM and LTM must be linked. If a participant is asked to explain each step and give reasons for

automatic processes that would normally not be attended to, interference can be expected since the participant must make an additional search for that information. Jourdenais (2001) distinguishes this type of verbalization from thinking aloud and calls it “introspection.”

Ericsson and Simon (1993) point out that the pure concurrent verbalizations of level 1 may be incoherent and disjointed, but they maintain that it is necessary to forgo coherence and completeness in order to study the unchanged cognitive processes involved in performing a task. Asking participants for explanations and verbal descriptions might not result in a representation of normal online thinking. In their words:

It is important to note that subjects verbalizing their thoughts while performing a task do *not* describe or explain what they are doing – they simply verbalize the information they attend to while generating the answer. When subjects verbalize directly only the thoughts entering their attention as part of performing the task, the sequence of thoughts is not changed by the added instruction to think aloud. However, if subjects are also instructed to describe or explain their thoughts, additional thoughts and information have to be accessed to produce these auxiliary descriptions and explanations. As a result, the sequence of thoughts is changed, because the subjects must attend to information not normally needed to perform the task (p. xiii).

Thus, participants can verbalize the thoughts that enter their attention and even *maintain* them in attention until they finish their verbalizations. The crucial point for nonreactivity is that the sequence of thoughts must remain the same with the production of a verbal protocol as it would be without it.

2.3.1.2 Nonveridicality

As we have seen, one of the core components of Ericsson and Simon's theory is the idea that concurrent verbalization can correspond to the information that is being attended to in STM during performance of a task. In order to make their explanation non-circular, though, it is necessary to know the contents of STM based on independent evidence (Brinkman, 1993). Participants' behavior must also corroborate their verbalizations in some way (Smagorinsky, 1989). Brinkman (1993) remarked that many studies utilizing protocol data have had methodological imperfections threatening generalizability in that they have tested only for reactivity or for nonveridicality, but not for both. With either kind of omission, he asserted, "true deviations from verbal report accuracy [may] go unnoticed" (p. 1383).

Naturally, there is a hierarchy according to which certain deviations are more serious than others. Some researchers have pointed out that the question of reactivity must take precedence over that of nonveridicality (Russo, Johnson, & Stephens, 1989). It seems to make sense that there is little point in testing whether or not participants are reporting processes accurately if the processes themselves have been altered by the condition of verbalization. However, if nonreactivity can be assumed, then it is important to check whether or not participants are committing "errors of omission" by not reporting

some of their thoughts or, alternatively, “errors of commission” by reporting mental events that do not actually occur (Russo, Johnson, & Stephens, 1989). In order to do this, one can compare verbal protocol data with data from another elicitation procedure, using process-tracing performance data alongside a concurrent report (Faerch & Kasper, 1987). In fact, Russo et al. have stated that it may not be possible to test veridicality without having recourse to this kind of simultaneous measure.

In Ericsson and Simon’s opinion, while measures of eye movements alone are not “fully adequate for catching the fine grain of thought processes,” they can be used redundantly as a means of validating verbal reports (1987, p. 51). The veridicality of participants’ verbalizations can also be assessed with the help of computer simulations that generate acceptable a priori models of information processing and regenerate observations (Faerch & Kasper, 1987; Ericsson & Simon, 1987). Since studies by Williams and Davids (1997) and Brinkman (1993) have checked for veridicality using verbal protocols along with eye movements and computer models, respectively, it may be instructive to examine them in detail.

Williams and Davids performed an experimental study with soccer players, asking them to watch videos of soccer simulations, verbalize where they were directing their attention visually, and also verbalize as quickly as possible the final destinations of passes. Their eye fixations were recorded and compared to what they said in order to assess eye movement and verbal protocol data as measures of selective attention. In the words of Williams and Davids:

Since verbal reports provide a direct measure of attentional allocation, the aim was to examine the association between visual orientation (as implied from eye-fixation data) and visual attention. If a meaningful relationship were demonstrated, this would support the validity of using either method as a measure of selective attention in human performance research (p. 366).

As far as reactivity is concerned, verbalization had no effect on performance in 11-on-11 soccer situations, and the participants were able to report accurately where they were looking. That is to say, the verbal reports and eye-fixation data did not contradict each other. According to Williams and Davids, the foveal vision employed in this kind of task was conscious and thus easily accessible for report. In 3-on-3 situations, on the other hand, there was reactivity (a slowing-down effect and a detrimental effect on task performance) related to the requirement to verbalize information about peripheral vision. In view of that, Williams and Davids suggested that since peripheral vision is usually unconscious, talking about it might disrupt task automaticity, leading to reactivity.

As for veridicality, Williams and Davids found that the accuracy of the two methods depended on the nature of the stimulus. When peripheral vision was necessary to complete the task, verbal reports were more valid than eye fixations because the participants could provide more information about their attention. However, when the task simply required foveal vision and the participants needed to change the location of their attention rapidly, eye fixations were more accurate, presumably since it was difficult to keep up in speech with the rapid changes. Their conclusion was that one should not

rely solely on eye-movement data when investigating visual search strategies and selective attention including peripheral vision. Instead, the two types of process-tracing data (verbal reports and eye movements) can be combined effectively.

Williams and Davids' conclusions make sense, and the fact that eye movements and verbal reports largely seem to corroborate each other may be useful for applications to other visual search studies involving selective attention. However, SLA researchers must keep in mind that their findings may not be generalizable to other kinds of tasks or other kinds of verbalizations. In Williams and Davids' study, the participants were asked to verbalize very specific information regarding where they were looking and what they were paying attention to – not what they were thinking. It is very likely that their reports did not represent all of their thoughts, and it is also questionable whether anyone would actually *think* something along the lines of, "I'm looking to the left now." This is more like a meta-description of one's actions, and it is probably not the kind of information SLA researchers are interested in, especially if they wish to follow Ericsson and Simon's (1993) guidelines against reporting on actions. It should also be noted that even on such a seemingly simple task, the techniques were not, in fact, found to be completely equivalent. Instead, the degree of mutual validation depended on the nature of the task. In SLA research, checking for reactivity and nonveridicality would be even more complicated than this.

Brinkman (1993) performed an experimental study to investigate whether or not validity and non-reactivity could be achieved using concurrent and retrospective verbal protocols on a fault-diagnosis task. The task he used involved looking at graphically-displayed networks consisting of rows and columns of components that were

interconnected in a variety of ways. The participants had to check the connections between the components to figure out which one did not work, and they carried out their analyses under three conditions: silent, concurrent verbalization, and retrospective verbalization. Brinkman decided to investigate reactivity because of the possibility that concurrent verbalization would induce a participant to perform the task using strategies that were easier to talk about. He also felt it was necessary to check veridicality, reasoning that the verbalizations might not be able to keep up with the speed of certain automatic recognition processes and therefore might not provide a complete picture of the strategies used. Both of these issues are clearly relevant to SLA research, especially since participants must often verbalize their thoughts in their L2.

Brinkman believed that a computerized fault diagnosis task would lend itself well to finding deviations from verbal report accuracy because it would be possible to use a computer algorithm to infer strategies based on the tests a participant made. Two basic (idealized) strategies could be identified: tracing-back (TB) and hypothesis-and-test (HT), along with another which could be labelled indefinite (IN). With a TB strategy, a participant would not pay attention to whether or not his tests were redundant; he would just test many times quickly until he found the faulty component by trial and error. Using HT, on the other hand, a participant would formulate a plan and then perform each test within that plan. Brinkman made note of the amount of time taken to complete each problem as well as the number of test trials used.

Veridicality was then checked by comparing human raters' strategy codings of verbalizations (protocol data) with strategy codings made by the computer algorithm (performance data), and a moderate degree of agreement was found. The concurrent

think-aloud condition slowed down the process (as could be expected from a level 2 verbalization involving putting information into an oral code), but it did not affect accuracy or strategy, and Brinkman was able to conclude that the strategy-related data from the concurrent verbalizations were more valid than those from the retrospective condition. Even though concurrent verbalization caused mild reactivity in the sense that it slowed down the process, he did not see this as critical. In discussing the seriousness of various kinds of invalidities, Russo et al. (1989) have declared that “disruption of the primary process is unacceptable, omissions in the verbal report are less serious, and a prolonged response time is usually inconsequential” (p. 767).

To sum up, then, both Williams and Davids (1997) and Brinkman (1993) found that concurrent verbalizations can be veridical, with the important caveat that the accuracy of the method may depend on the nature of the stimulus or task. That makes it especially important to consider whether the kinds of tasks used in their studies are comparable to the kinds used in SLA research. In Brinkman’s research, computer algorithms could be written, and two major strategies could be identified for discrete, finite problem solving. In studies of writing processes, on the other hand, participants cannot simply use trial-and-error or hypothesis-testing strategies with the benefit of immediate feedback. With this in mind, SLA researchers should be sure to check for nonveridicality and reactivity in their own studies. In some writing tasks, it may be possible to use eye-movement data or stimulated recalls to compare silent participants with participants who perform concurrent verbalizations. But in any case, SLA researchers cannot take it for granted, based on research in other fields, that the use of verbal protocols will not interfere with the processes they are trying to study.

2.3.2 *Effects of verbalizations with different kinds of tasks*

Ericsson and Simon (1993) described various kinds of cognitive tasks, such as anagrams, geometry proofs, and number puzzles, implying that with the proper instructions and the restriction of reporting only on information in STM, verbalization should not interfere with cognition (Russo et al., 1989). Although the studies discussed above certainly have limitations, they do seem more or less to support this.

However, some researchers have asserted that not enough direct studies of reactivity have been undertaken to predict exactly when concurrent verbalization will interfere, and that certain results of their experiments contradict what Ericsson and Simon's theory would predict. Russo, Johnson, and Stephens (1989) maintained that the theory was not, in fact, adequate to indicate *a priori* which kinds of tasks would be affected by verbalization, while Stratman and Hamp-Lyons (1994) have additionally pointed out that when reactivity is present, it is not always clear what the effects can be attributed to or in which direction they will go. For example, they ask, will the requirement to think aloud hurt novices because of STM demands, or help them because they have to verbalize reasons for what they are doing? Will it hurt experts because they have to verbalize processes that are normally automatic? If experts provide fewer concurrent verbalizations than novices, are they experiencing greater or less interference from the verbalization requirement? With questions similar to these in mind, Russo et al. stated that researchers must verify nonreactivity empirically (i.e., by adding a silent control group and looking at accuracy and response time) until a theory of verbal protocols can state definitively what the conditions of validity are.

Russo, Johnson, and Stephens (1989) proposed that the causes of reactivity are not general; rather, they are created by the combination of the task and verbalization demands. Although Russo et al. believed that it would be difficult to know beforehand whether or not reactivity would occur, they chose four different problem solving tasks for their study: anagrams (verbal in nature), gambles (numerical, involving very simple mental multiplication), Raven's matrices (pictorial/visual), and mental addition (carrying a heavy STM load). These were chosen because the first two apparently satisfied Ericsson and Simon's conditions, while the other two violated them. For the anagrams and gambles, no reactivity was expected. On the other hand, it was reasoned that reactivity could be expected for the addition and Raven's matrices tasks because of the importance of mental rehearsal of partial results and the recoding from a pictorial to an oral code, respectively.

To test these hypotheses, Russo et al. compared participants solving the problems silently to participants who solved them while thinking aloud, and they also looked at eye fixations to monitor the addition task. In accordance with their initial expectations, it was found that verbalization made participants significantly less accurate on the addition task while having no effect on anagrams. However, even though the gambles task was not predicted to show reactivity, the participants were significantly more accurate in the verbalization condition, and while reactivity *was* expected for the Raven's matrices, none was found for either time or accuracy. Thus, the predictions that Russo et al. had made based on Ericsson and Simon's theory did not fit well with the data they obtained, and they concluded that reactivity depends on the task. Still, believing that verbal protocols are an extremely useful research tool despite the unpredictability of reactivity, they called

for more research to be done to identify the best ways to reduce serious invalidities and identify the causes of reactivity.

Stratman and Hamp-Lyons (1994) point out that many writing researchers assume it is practically impossible to predict or discern the effects of think-alouds since they may be highly individualized or even random. They are optimistic in this regard, however, and note that it may be possible to refine Ericsson and Simon's theory and find systematicity to the interference of verbal protocols, even with respect to writing tasks, which are considered to be less well-defined than others. Since the oral compatibility of STM contents is apparently not the only key to whether or not a think-aloud protocol may be reactive, writing researchers should try to figure out the reasons for the various effects that verbalization has on different tasks. Specifically, Stratman and Hamp-Lyons hypothesize that producing a think-aloud might help learners notice more surface errors in their writing, but it might hinder them from detecting large organization problems because of the limitations of short-term memory.

As one step in this direction, Stratman and Hamp-Lyons (1994) carried out a pilot experimental study comparing think-aloud and silent conditions on a revising task with isomorphic comparison-contrast paragraphs. They gave their participants as much time as they needed and looked at three different output measures to see whether they were affected by the requirement of verbalization. They wanted to find out if thinking aloud would affect the participants' ability to 1.) detect errors and produce acceptable revisions, 2.) avoid introducing new errors, and 3.) preserve meaning or introduce new content. They found that the think-aloud condition was associated with a lower ability to detect organization errors, a higher ability to detect faulty pronoun references (possibly because

of acoustic feedback), and the introduction of twice as many new word-level errors and new sentences. The silent condition was associated with more meaning changes with word or phrase additions, deletions and substitutions. Of course, being a pilot study, the results are not conclusive, but they do show clearly that think-alouds may enhance the revision process in some ways and hinder it in others.

2.3.3 Factors causing reactivity

As other researchers have done, Stratman and Hamp-Lyons attempted to delineate five factors associated with the use of concurrent protocols that may cause reactivity, finding the following:

- 1.) experimental task directions... that elicit an inappropriate level of verbalization,
- 2.) limited STM capacity for talking and attending at the same time,
- 3.) hearing one's own voice,
- 4.) learning that occurs because thinking out loud increases subjects' critical attention to their activities, and
- 5.) direct or indirect experimenter influence through verbal or nonverbal cues (p. 95).

These factors can be compared to those outlined by Russo et al. in their 1989 study, all, of course, independent and task-specific:

- 1.) the attentional demand for processing resources (corresponding to Stratman and Hamp-Lyons' number 2),
- 2.) auditory feedback, which can either facilitate or interfere with performance (corresponding to number 3 above),
- 3.) enhanced learning over repeated trials, and
- 4.) a motivational shift towards greater accuracy.

Jourdenais (2001) also mentions concerns that overlap with those listed above, such as memory constraints, the desire to please the researcher, the effects of elicitation techniques, extra learning opportunities that may occur, and the question of whether or not participants have the metalinguistic knowledge to be able to describe their behaviors.

It would be beyond the scope of this paper to go through each of these in turn. However, in considering how verbal protocols can be applied to L2 research, there are certain reactivity-causing factors that seem particularly relevant. Most of the experimental studies that have been discussed so far have dealt with discrete problem-solving tasks. Since L2 research has its own special characteristics, it will be important to consider the distinction between declarative and procedural knowledge, the kinds of L2 tasks that are studied, various practical concerns involving the presence of the experimenter and the task instructions, and the limitations and cognitive load involved in having to verbalize all of one's thoughts in an L2.

2.3.4 Applicability to L2 research

As has already been mentioned, more and more L2 researchers have been using verbal protocols despite questions that have been raised about the methodology (Cohen, 1987). For example, Grotjahn (1987) has questioned whether Ericsson and Simon's ideas can be applied directly to SLA research since it is not (or at least not purely) an investigation of problem solving. Others have wondered whether learners' verbalizations truly represent internal reality or if introspection might instead involve people's hypotheses about what must be happening, based on implicit theories or rules of thumb that they have developed – in other words, what people think they know (Nisbett & Wilson, 1977; Seliger, 1983, as cited in Gass & Mackey, 2000). Grotjahn has also inquired about the ontological status of interlanguage, asking whether it is a state of mind that can be accessed or merely a theoretical construct that is not actually instantiated. According to Cohen (1987), it may be problematic that when language learners are asked to verbalize about the processes they are going through, language has to serve two functions: task performance and process description. Cohen has also pointed out that, especially with language, it may be difficult to tell whether participants are actually thinking aloud (without analyzing their thoughts) or whether they are, in fact, observing their thoughts on another level and reporting on those observations.

Some researchers have brought up the distinction between declarative and procedural knowledge, with *declarative* knowledge referring to language learners' analyzed and organized knowledge of rules, and *procedural* knowledge referring to the mostly automatic cognitive and interactional processes involved in language reception, production, and acquisition (Faerch & Kasper, 1987; Gass & Mackey, 2000). Procedural

knowledge is said to intervene between declarative knowledge and linguistic behavior, activating declarative knowledge in communication and extending it through learning. However, since it is mostly automatic, it is not maintained in STM and is not available for report. Declarative knowledge, on the other hand, can be accessed directly and verbalized (Gass & Mackey, 2000).

As far as protocol data are concerned, then, if conscious attention to a process is not necessary, no insights regarding it will show up in verbalizations (Dechert, 1987). Dechert (1987) has suggested that in a translation task, for example, some of what does not show up in a verbal protocol may be related to the automatic recognition processes involved in processing the text. Nevertheless, Faerch and Kasper (1987) have noted that if a participant experiences a breakdown of an automatic process, he or she will pay attention, and then the processes used will be available for report. They have also stated that it may be possible to introspect on and verbalize about procedural knowledge during certain slow and controlled activities, such as written translation.

Some of these points would seem to be true no matter what kind of task is being considered. For some activities, conscious attention is required and sequences of steps can be broken down into parts, whereas for other activities, it would be difficult or strange to think about the steps involved. Whether it is an L2 task or a motor task like driving a car, if a process involves automatic recognition, it will not show up in the protocol data. The distinction between declarative and procedural knowledge is not unique to language use. Still, since language tasks may require that everything take place through a linguistic channel, the input and output might interfere with each other, and

reactivity and nonveridicality might occur differently from how they appear in other kinds of studies and tasks.

2.3.5 Task characteristics

Some SLA researchers have specifically called attention to differences between the kinds of tasks and verbal reports described in the cognitive science literature and those used in the analysis of language processing (e.g., Dechert, 1987). In much of the literature discussing the implementation of think-aloud protocols, the focus is usually on well-defined, achievable, sequential tasks with specific goals and identifiable end-products. Often, experimenters already know about the inherent structure, rules, sequences of steps, and strategies associated with certain kinds of problems and their solutions (Dechert, 1987). This allows researchers to check for accuracy and compare participants with each other (Stratman & Hamp-Lyons, 1994). In “ill-defined” language tasks, though, participants may have their own goals, and there may be many acceptable solutions to any problem. It is sometimes difficult to tell if the requirement to verbalize is interfering with a task because no one can say what an “accurate model” looks like (Stratman & Hamp-Lyons, 1994).

The very fact that there may be no clearly identifiable final goal may make certain L2 tasks qualitatively different (from the perspectives of both the participant and the researcher) from the other sorts of tasks discussed in the literature. In a math problem, strategies and sequential steps lead toward a solution that can be expressed as a single number, and the participant is aware of this. However, in a revision task, for example, the process of revising does not necessarily unfold in logical, sequential, directional

steps; it might be difficult to tell how much detail to go into, how much time to spend on various parts of the task, and when the task is finished. In other words, the overall “problem” is not simply worked out in a series of steps that lead toward an easily expressible ultimate solution; rather, each instance of noticing something in the original essay might be considered a separate problem with its own steps to understanding it. Moreover, it might overlap with or be related to other instances of noticing.

Basically, a lack of structure may make it so that not all participants mentally construct a task in the same way. As we will see later, a participant may try to perform a task in the way that he or she assumes the experimenter expects, and the requirement to talk aloud may encourage the participant to think about things just so that he or she can talk about them. From the point of view of the participant, then, the minimal task is not just to revise (and, incidentally, also to speak thoughts out loud), but rather to *have things to say* about revising. According to Jourdenais (2001), the production of a think-aloud protocol is an extra task.

2.3.6 Training and instructions

Even though the main task under investigation may be ill-defined, some L2 researchers have argued that training can help to make at least the process of producing a think-aloud more well-defined. It has been observed, for example, that participants who are simply asked to think aloud while reading tend to read long passages of text and then retrospect on what they have read. Training them instead to verbalize whenever they pause while reading can help to avoid this (Cohen, 1987). This does not make the task or processes of reading better defined, but it can affect the task of verbalization, which

researchers have assumed might then be more effective in bringing the processes of reading to light. Ericsson and Simon (1993) claimed that training does not affect the validity of verbal reports; it only has the effect of increasing the completeness of verbalization. However, since some L2 researchers have asserted that training may bias learners to verbalize certain things, this claim should be investigated empirically with regard to L2 tasks in particular (Jourdenais, 2001; Faerch & Kasper, 1987; Gass & Mackey, 2000).

According to Gass and Mackey (2000), if pilot studies show that participants need training, then they should be trained to the point that they are able to carry out the procedure. Nevertheless, it is important that this training remain minimal to avoid letting the participant in on experimental goals or unnecessary information. It is also crucial that instructions be standardized. Since even minimal differences in instructions can affect the nature of a participant's verbalizations, participants should be given exactly the same instructions, whether this means recording them, reading them from a script, or presenting them to the participants in written format (Gass & Mackey, 2000).

Ericsson and Simon explain that instructions to think aloud are usually very short and simply make reference to an activity that the participants are presumed to be familiar with (1987). For example, suggesting that the thoughts would already have the form of inner speech, Duncker (1926) would tell his participants, "Try to think aloud. I guess you often do so when you are alone and working on a problem." Claparede (1934) would say, "Think, reason in a loud voice, tell me everything that passes through your head during your work searching for the solution to the problem." Since this refers to everything in the person's mind, a participant might have to recode some information into verbal form.

Krutetskii (1976) went a little further and mentioned the importance of not trying to explain thoughts to anyone else. He said, “Pretend there is no one here but yourself. Do not tell about the solution but solve it” (all as cited in Ericsson & Simon, 1987, p. 36).

A more modern and elaborated form of instructions can be found in Steiner (1986), who has suggested saying the following:

- 1.) Say whatever’s on your mind. Don’t hold back hunches, guesses, wild ideas, images, intentions.
- 2.) Speak as continuously as possible. Say something at least once every 5 seconds, even if only, “I’m drawing a blank.”
- 3.) Speak audibly...
- 4.) Speak as telegraphically as you please. Don’t worry about complete sentences and eloquence.
- 5.) Don’t overexplain or justify. Analyze no more than you would normally.
- 6.) Don’t elaborate past events. Get into the pattern of saying what you’re thinking now, not of thinking for a while and then describing your thoughts (p. 701).

Russo, Johnson, and Stephens (1989) have also emphasized the importance of encouraging participants to be more concerned with naturalness than with completeness.

As a matter of fact, when utilizing verbal protocols as a research tool, one of the most important parts of the instructions might be a warning against “self-theorizing or

other introspective explanations” (Russo et al., 1989, p. 759). First of all, it has been observed that when instructions do not request motives and reasons, participants do not include them in protocols, suggesting that the reasons may not normally be conscious (Smagorinsky, 1989). Furthermore, Nisbett and Wilson (1977) contended that participants are often inaccurate when trying to rationalize about their reasons for doing things and tend to hypothesize about processes instead of reporting their actual thoughts. In other words, participants may draw on their own implicit, a priori theories instead of making reference to actual thought processes – although it should be noted that this might be more of an argument against the use of stimulated recall, which is done following the completion of a task, than against concurrent verbalization. In any case, reporting sequences of thoughts is different from trying to give reasons for a thought sequence (Ericsson & Simon, 1987). Ericsson and Simon have noted the danger of encouraging participants to ask themselves, “What am I doing now?” since, presumably, the more they try to come up with descriptive terms in order to report on their activities, the more their normal underlying cognitive processes may change (Stratman and Hamp-Lyons, 1994).

Hayes and Flower (1983) provided a hypothetical example of how task performance might be affected by the requirement to talk about something that would not normally be heeded. If an experimenter asked a participant to divide two large numbers in his head and talk aloud, but also mention every time he noticed an odd number, the verbalization might go something like this: “248 into 1336 is about 5, so 5 times 248 is – oops, 5 is an odd number – now where was I? Is there something important about odd numbers in this problem? Oh, yeah, 5 – that’s an odd number – well...” (p. 215). A real example can be found in Toms (1992), a study using the same fault diagnosis task as the

one used by Brinkman (1993) in the study described earlier. In contrast to Brinkman, Toms found that concurrent verbalizations not only slowed down processing but also caused impaired accuracy on the task. Brinkman explained these inconsistent results by suggesting that they might have had to do with the way in which Toms tried to elicit the verbalizations. At certain moments during the task, Toms encouraged the participants to report on very specific information. This could have disrupted their performance because the requested information might not normally have come into the participants' attention (Brinkman, 1993). Another real example comes from Gagne and Smith (1962), who purposely used different sets of instructions, with one set requesting reasons for each move the participants made. According to Gagne and Smith, the fact that the participants had to think more about the processes in the reason-giving condition may have been the source of their better performance (as reported in Smagorinsky, 1989). These examples make it clear that being instructed to explain the steps of a solution can be very different from focusing all of one's attention on solving a problem efficiently while verbalizing concurrently (Ericsson and Simon, 1993).

2.3.7 Experimenter influence: Social interaction

Besides the explicit instructions, there are many variables that can affect participants and influence the kind of information that makes its way into their verbalizations. For example, Smagorinsky (1989) mentions the conditions of the protocol situation, including the researcher's behavior and time constraints, while Cohen (1987) would add to that the number of participants, the mode of elicitation and response, whether or not the situation is videotaped, how the instructions are given, and how much

formal structure is imposed by the researcher. Participants may naturally feel that they have to make themselves intelligible or articulate things that are partially automatic because of the presence of a researcher (Russo et al., 1989). Moreover, even if participants initially seem to understand the instructions, it may be easy to lapse into default mode since explanations are such a familiar form of verbal communication (Ericsson & Simon, 1987).

Interactions between the participant and researcher can have a considerable impact on the data, even if the verbalization superficially seems to take the form of a monologue, without any feedback (Faerch & Kasper, 1987). The experimenter has to be extremely careful that he or she appears to be more like a “warm body” (or even nonexistent) and less like a conversation partner (Gass & Mackey, 2000, p. 60). This can be partially accomplished by sitting out of sight behind the participant to make it clear that social interaction is not intended. A researcher should try to be as neutral and unobtrusive as possible (Smagorinsky, 1994).

Even if these rules are followed, though, the observer’s paradox still applies. The simple fact of being observed can change a process, and there are many inherent human characteristics that can affect participants’ behavior. For instance, male and female experimenters have been shown to obtain different results from participants, and males and female participants may also elicit different kinds of behavior from researchers (e.g., how much they smile, how attentive they are, how much friendliness and warmth they show, etc.). Other characteristics to consider are the participant’s age, need for approval, and acquaintance with the researcher, and the researcher’s age and expectations. A researcher can cue desired behaviors completely unintentionally.

Participants often try to be as cooperative as they can. As mentioned earlier, instructions and training might give them a clue as to what the researcher's expectations are, and that might have an effect on the kind and amount of information they report (Hayes & Flower, 1983). The presence of an experimenter can also cause a motivational shift in that if participants know that their errors will be somewhat "public," they might use strategies that reduce errors but require more effort than normal. According to Russo et al. (1989), they may also try to act in accordance with what they think the experimenter prefers. If, for example, the researcher also happens to be the participants' teacher, they may seek approval or try to display knowledge of things the teacher has mentioned in class.

In order to ensure that participants speak continuously in a think-aloud protocol, it is sometimes necessary for the experimenter to prompt them. Such prompting should be kept to a minimum, and it should be nondirective and standardized (Russo et al., 1989). Saying something like "Keep talking" is better than saying "Tell me what you are thinking," which might be perceived as a social request, or "What are you thinking about?" which is more likely to encourage the participant to engage in self-observation or produce an "other-oriented" description (Ericsson & Simon, 1987, 1993).

If all the warnings mentioned above are heeded, it is often assumed that a distinction between pure concurrent verbalization and social verbalization can be maintained. Ericsson and Simon have argued explicitly that concurrent verbalizations can be isolated from interactive uses of language. However, according to Hauser (2002), this distinction is impossible in practice. Hauser performed an experiment in which his participants worked on a computer program targeting the use of the definite article with

proper names. He elicited concurrent protocols from them during the exposure phase and also had a retrospective “post-experiment judgment” interview about their behavior. He found that even though in the concurrent verbalizations there were no indications that the learners had been using intentional learning strategies, some of them mentioned in the retrospective interviews that they had been looking for rules. This could be corroborated by the fact that the participants who mentioned looking for rules performed more accurately than those who never mentioned looking for rules at all.

As far as Hauser could tell, the conditions seemed very conducive to eliciting non-social verbalizations. The experimenter was the only other person in the room, he was seated several feet behind the participant and could not be seen, and he never spoke during the concurrent verbalization. Furthermore, the participants seemed to understand the directions, one of them saying, “So, if I think now I’m hungry, so I say I’m hungry...” At first, the protocol data seemed to fit the description of a Type 2 verbalization. Nevertheless, upon closer inspection, interactive uses of language were evident. For example: “It never snows on Mount Fuji.... no way... every winter uh top of the mountain with covered with snow and white snow and blue mountain is very beautiful.... *uh what shall I say? Yeah anyway, so I like Mount Fuji very much....*” These statements were not necessary for the completion of the task; rather, the participant searched for relevant comments to make, and thoughts (personal experiences and opinions) entered his mind because of the way in which he had assessed and constructed the task. Hauser concluded that the participants probably made mention of noticing only if they thought that such reporting was relevant to the task. He also asserted that all verbalizations may be Type 3 (inappropriate and reactive) since participants necessarily

search for and select specific types of information for report instead of just relating the contents of STM. L2 researchers cannot assume that verbalization merely affects the amount of time participants take to complete tasks.

To sum up, then, Hauser found both reactivity and nonveridicality in his study. His participants talked about topics they would not normally have mentioned to themselves in the process of completing the task, and they also did not talk about all of the things that they were actually thinking. It is possible to speculate about how Hauser's argument that all verbalizations are Type 3 makes sense for SLA research. As we have seen, whereas Ericsson and Simon discussed many studies dealing with small, well-defined tasks that could be completed one after the other (a series of math problems, for example), writing and other L2 tasks may be different in that a participant does not simply go through a series of sequential steps to solve one small problem and then move on to the next. Brinkman (1993) and Williams and Davids (1997) addressed important issues of reactivity and validity in their studies, but their studies included tasks that were different from language tasks with regard to concreteness and the sorts of things that could be verbalized (e.g., simple trial and error processes). Even Russo, Johnson, and Stephens (1989), who used tasks involving words, numbers, images, and high STM demands, nevertheless used more or less discrete, finite problems with solutions. Language tasks are often ill-defined, and because of the nature of language, some may inherently entail an intention to communicate. In addition, the type of knowledge that L2 learners tap into might be more or less accessible to introspection, and they might be more or less inclined to engage in meta-analysis of their actions (i.e., descriptions, explanations, reports).

Russo et al. (1989) proposed that the reactivity they found in their study occurred as a result of the combination of task demands and verbalization. Interestingly, on their gambles task, for which no reactivity was expected, verbalization produced a positive effect. On their mental addition task, for which reactivity was expected, verbalization produced a negative effect. The question is not whether or not verbalizations can provide more insights regarding processes inside learners' heads; in fact, it seems certain that they are useful for that purpose. What is in question is how concurrent verbalizations can be used as a nonreactive and veridical (L2) research methodology, that is, one that accurately represents what is going on in learners' heads and does not change the thoughts they would normally have while performing a certain kind of task.

2.3.8 Verbal protocols in an L2

On top of what has already been mentioned, it is also clear that speaking in an L2 while thinking aloud can place additional demands on STM and affect the cognitive processes involved in completing a task. Depending on proficiency, it may also affect the sorts of thoughts a learner is able to express. If in an experimental study the participants speak more when using their L1, this might provide evidence that they are not expressing everything that they are thinking when they use their L2. Alternatively, it might suggest that using an L2 actually hinders or changes thought processes in some way. On the other hand, it is also possible that using an L1 while trying to discuss an L2 could interfere with language processing.

The physical act of simply producing an utterance is assumed not to affect cognitive processes (Smagorinsky, 1989), and in Brinkman's fault diagnosis study, he

stated that while verbal recoding does put some demands on STM, “as long as there are *verbal codes available* which make the recoding fairly easy, the course and structure of the processes should not be affected” (1993, p. 1394, emphasis added). According to Stratman and Hamp-Lyons (1994), there is an assumption in reading and writing research that it is relatively easy to verbalize the contents of STM during a reading or writing task since thoughts do not have to be recoded. However, it may be quite problematic to try to apply these ideas to L2 tasks in particular. In an L2, coming up with the terms necessary to express thoughts is not a highly automated process; it requires considerably more effort and active search than it does in an L1. Without having the verbal codes available, L2 learners may not be able to talk as quickly as they can think; they might get stuck on one point and lose other important information, causing them not to be able to pursue a particular line of reasoning.

Russo et al. (1989) state that a primary task and verbalization may compete for processing resources; by extension, this may be especially true when both the task and the verbalization require the use of an L2. Participants must figure out how to allocate their resources. If they must maintain items in STM so that they can figure out how to talk about them, that might reduce their ability to focus on the primary task, which, involving language, may require a great deal of processing resources itself. If participants use fewer resources for the main task and more for the purpose of verbalization, that might cause reactivity. If they use more for the task and fewer for verbalization, that might cause nonveridicality. According to Russo et al., participants probably assess the relative costs of doing each in a particular task situation.

Gass and Mackey (2000) similarly point out that L2 learners might verbalize just what they feel they are able to express. In a study of learners of English as a Second Language (ESL) and Italian as a Foreign Language (IFL) who had to produce verbal protocol data in English, Mackey, Gass, and McDonough (2000) found that the average number of words per recall comment for the IFL learners (for whom English was a native language) was 26, whereas for the ESL learners (for whom English was obviously not a native language), it was only 16. The ESL learners may have had ideas that they wanted to express, but the constraints of their L2 may have made it more difficult to do so. This particular study employed stimulated recall; however, it is easy to see how these findings can be applied to concurrent verbalizations as well. It is important in both kinds of verbal protocols to assess the participants' ability to verbalize in an L2, especially with regard to the expected demands of a particular verbalization task. Gass and Mackey (2000) have stated that since some things may be easier for learners to verbalize than others, researchers should make use of pilot testing to check whether participants have the necessary linguistic competence. It should also be noted that learners activate many kinds of linguistic knowledge when carrying out L2 tasks, and how much of an impact any necessary recoding has on the way tasks are carried out remains an issue for future research (Faerch & Kasper, 1987).

Researchers should not forget that producing a think-aloud protocol, especially for L2 learners, may be equivalent to carrying out an additional task. Even in an L1, think-alouds might mean a greater cognitive load, but requiring learners to verbalize all of their thoughts in an L2 may cause the process to be very different from what would happen if they could focus all of their attention simply on performing the task efficiently.

It is also important to consider that L2 learners may be less willing to take risks with language and may be more worried about making ungrammatical utterances, possibly avoiding linguistic items that involve IL gaps (Jourdenais, 2001). Cohen and Olshtain (1993) remarked that L2 learners may differ in their production styles and levels of comfort with their own output; whereas pragmatists might care most about simply being understood, avoiders might utilize circumlocution so that they do not have to use certain structures, and metacognizers might focus on monitoring their grammar and pronunciation. These styles can clearly have an effect on learners' verbalizations.

2.4 Summary

This thesis being an attempt to investigate L2 learners' processing of written feedback, quality of noticing, and the relationship between noticing and subsequent language production, three bodies of research seem particularly relevant: research suggesting that reformulations might be able to address some of the problems associated with explicit error correction, research on noticing, and research on the reactivity and nonveridicality of verbal protocol data. Many researchers agree that L2 learners need both positive and negative evidence for SLA, and it seems as though writing should be a useful medium for the provision of corrective feedback, especially considering that written language is less fleeting than spoken and provides concrete opportunities for learners to focus on form and meaning. Teachers and researchers have argued, however, that explicit error correction might not be worth the time and effort, given its many practical problems and doubts about its effectiveness. If noticing truly is necessary for

converting input to intake, then it is important to consider the amount and quality of noticing that learners experience. It is also important to realize that this can be affected by the perceptual salience of forms or corrections, the learners' skills, the task demands, the amount of automaticity involved, and many other factors.

SLA researchers have emphasized the value of noticing both similarities and differences between IL and TL, that is, realizing what one is able to produce and what one is not yet able to produce (i.e., "noticing the gap"). With this in mind, reformulation has been proposed as an alternative to error correction. The idea is that if students have to search for differences, process them deeply, and evaluate them when looking at reformulations, they might be engaging their IL systems to a greater extent than often seems to occur with other kinds of corrective feedback. This high involvement load may be helpful in promoting noticing and quality of noticing. Theoretically, it seems that reformulation should involve both error analysis and cognitive comparison; it should also provide both negative and positive evidence to the effect that students have the opportunity not only to recognize what may be prohibited in the target language, but also to acquire new language by receiving comprehensible input at higher levels of sophistication and complexity. Since learners must focus on meaning and form in order to make sure reformulations express what they have intended, this kind of feedback might also be processed more deeply. Students might develop more cognitive strategies for noticing and increase their levels of awareness about their own common mistakes when they actively notice the same differences multiple times. Furthermore, when feedback is related to what students have already attempted to produce and will have to produce again in a revision, it might be more effective.

The use of think-aloud protocols may be able to help researchers investigate what L2 learners notice and whether some kinds of noticing are more effective than others. Besides simply using it as a hopefully nonreactive research methodology, though, it is possible that asking participants to verbalize in certain ways might either hinder or encourage more substantive kinds of noticing. Researchers have often focused on the reactivity of verbal protocols in a negative sense, and there are certainly ways in which verbalization can have a negative impact on task completion. But perhaps when applied to a writing/noticing task, talking aloud can promote increased attention, deeper processing, more reasoning, and ultimately better revisions. Since many factors have been proposed as possible causes of both positive and negative reactivity, reviewing them will help us later to understand the results of our study.

A common warning given with respect to verbal protocols is that training should be minimal to avoid influencing thought processes or giving the participants hints as to the nature of the research. Also, since explanations, reasons, and procedural knowledge may not normally be conscious, and since they might entail an additional internal search for relevant information to report, it is often advised that a researcher should not explicitly ask for them in the task instructions. Participants should not be encouraged to describe what they are doing in view of the fact that coming up with the terms to describe their actions could change the underlying thought processes. Explaining a problem step by step is different from simply solving it and talking at the same time. If participants have to make links between information in STM and LTM, it will not only slow down the process, but could also change it. Although this is undesirable when think-aloud

protocols are being used as a research methodology, we hypothesize in our study that it might actually enhance the kinds of linguistic noticing that take place.

Researchers have also cautioned that, especially in L2 research, verbalization may act as an additional task. People have limited STM capacity for talking and paying attention at the same time, and depending on L2 proficiency, the use of an L2 could limit this capacity even more. If speaking in the L2 is not a highly automated process, this might constrain the ability to verbalize, take attention away from the main task, and cause participants to lose information while they are trying to figure out how to verbalize their thoughts. Participants may also be less willing to take risks in an L2 and might choose not to verbalize certain thoughts if they do not know how to express them. These factors would presumably have negative effects on noticing.

There are factors besides the oral compatibility of the contents of STM that might affect noticing and revising as well. Simple auditory feedback itself might increase participants' attention to their activities, and in a revision task it might help them to notice more surface errors but fewer organizational problems, for example. The nature of a task and the way in which a participant conceptualizes the requirements of the task are also important. If an L2 task is relatively ill-defined, this lack of structure might cause participants to conceive of the task in different ways; they may search for and select different sorts of things as relevant to talk about. Moreover, even if participants understand the directions and know that they should just speak their thoughts out loud (incomplete or not), it may not be possible to avoid social interaction completely; participants may naturally fall into modes of conversation with which they are more familiar, speaking coherently for the benefit of the researcher who is present. Attempting

to keep the above issues related to noticing, reformulation, and thinking aloud in mind, we propose the following hypotheses for our research questions.

2.5 Hypotheses

RQ1: What do L2 learners notice as they compare their text to a reformulated version while thinking aloud? (corresponding to Qi and Lapkin's second research question)

H1: (Descriptive) We assume that the L2 learners in our study will notice a wide variety of errors, including lexical, morphological, and syntactic errors, as well as stylistic differences and errors of spelling and punctuation.

RQ2: How is such noticing related to changes in the written text completed after comparing the original and reformulated versions? (corresponding to Qi and Lapkin's third research question)

H2: Changes "noticed" will be associated with more corrections than those not noticed; changes "noticed" with a reason will be associated with more corrections than those learners do not give a reason for.

This hypothesis recalls and seeks to find support for Qi and Lapkin's (2001) finding that reformulation changes that were noticed with a verbalized reason were associated with more accurate revisions than those noticed without a reason. The

research in this thesis may be able to confirm quantitatively that there is a relationship between high quality noticing and the ability to revise later. As a matter of fact, research by Leow (1997) has also demonstrated through the use of verbal protocols that the different types of information learners provide in think-alouds may be related to linguistic accuracy on subsequent tasks. Leow found that learners who made metacomments, showed awareness at the level of understanding (not merely noticing), and stated rules about certain targeted forms performed more accurately on later tasks than learners who simply mentioned the forms without stating rules. Given that his research focused on the learners, the results could be related to their overall orientations toward language learning. However, further support comes from similar results in Leow (2003), showing that while simply noticing forms was helpful, demonstrations of higher levels of awareness with evidence of understanding were associated with the identification and use of target linguistic items.

When discussing these associations, it should not be overlooked that it is *not*, in fact, possible in this research to assert a cause-effect relationship between quality of noticing (itself) and subsequent linguistic accuracy. Quality of noticing may very well be influential or facilitative in some way, but it is also possible that learners demonstrate high quality noticing and explanations when they are ready to learn and use the particular structures that later show up in their revisions. Since learners' verbalizations may be evidence of their own developmental readiness, it is not possible to state definitively that the noticing itself causes what happens in the revisions, nor is it possible to define precisely what it is that makes students notice (or notice at a certain level). Still, even though a cause-effect relationship cannot be claimed, this thesis seeks to corroborate the

sorts of findings discussed above by investigating whether higher quality noticing may be associated with greater subsequent linguistic accuracy on a three-stage writing task involving reformulations.

RQ3: Do students notice more when comparing their essays to reformulated versions as opposed to versions with explicit error corrections?

H3: Comparing an essay to a reformulated version will lead to more noticing and more changes (i.e., greater linguistic accuracy) than simply looking at error corrections.

Qi and Lapkin assumed that effective corrective feedback not only encourages learners to pay attention to their errors, but also provides learners with more natural and sophisticated TL data so that they can notice the gap between IL and TL, based on their own interests and needs. If reformulations really do promote error analysis, cognitive comparison, active search and evaluation, and a more analytical orientation, we can ask if the correspondingly higher cognitive load might enhance or hinder noticing and subsequent correction. When we consider the ease of understanding and the aid of visual memory that may accompany a format like written error corrections, H3 might seem counterintuitive. However, it is possible that the involvement load aspects of need, search, and evaluation that have been shown to affect vocabulary acquisition might apply to corrective feedback as well. Perhaps the active comparison of a first draft with a native speaker's reformulation might lead to more rehearsal in STM, greater understanding, development of cognitive strategies for noticing, and retention of linguistic features than

occur with explicit error corrections – especially since in the latter case learners may simply be able to look at changes to their texts without much extra encouragement to process them deeply. If this is the case, and if the cognitive load is not too great, it would seem that learners in the reformulation condition should improve more on revisions than those in the error correction condition.

RQ4: Does the use of think-aloud protocols affect the number of linguistic features that students notice and that subsequently make their way into the final version of the written text?

H4: Thinking aloud while comparing an essay to a reformulation will lead to more noticing and changes (linguistic accuracy) than not thinking aloud.

According to Leow (2003)'s review of past research (Rosa & O'Neill (1999) in particular), L2 learners' level of awareness during a task appears to be correlated with the existence of formal instructions encouraging them to look for rules. If reformulations do encourage the sorts of approaches to feedback processing hypothesized above, then we can ask whether perhaps instructions to talk about the differences between two versions of writing might encourage them even more. Of course, knowing what we do about positive and negative reactivity, it should be stipulated that the learners must be of high enough L2 proficiency that the requirement to verbalize their thoughts does not disrupt the process too much. However, if speaking in an L2 is automated enough, the requirement to verbalize might induce learners to engage in further reflection and

problem solving. They might not only notice additional aspects of the reformulations, but they might even notice them at a higher level of understanding or a deeper level of processing than would occur with either explicit error corrections or the silent comparison with a reformulation alone.

Chapter 3

STUDY 1 (REPEATED MEASURES DESIGN)

In order to investigate these hypotheses, two separate but related studies were carried out. The first study, a repeated measures design, was conducted with 15 ESL learners. Then, in order to investigate the same questions while addressing some methodological issues, a non-repeated measures design was used with 54 participants. In the first study, each learner participated in three different writing conditions (error correction, reformulation, and think-aloud), counterbalanced to control for effects of writing topic and order of condition. In the second study, a control group was added, and each learner participated in only one of the four writing conditions. The students' essays and revisions in both studies were analyzed in order to compare changes in accuracy (as possible evidence of noticing) among the conditions.

3.1 Participants (Study 1: Repeated Measures)

The original participants in Study 1 were 31 high-intermediate ESL students in the Intensive English Program (IEP) at a large Midwestern university. However, due to absences and the desire to balance out the number of participants in terms of order of condition, only 15 of the participants' data were used for analysis. Of these fifteen, 11 were Korean, 3 Japanese, and 1 Indonesian. The female to male ratio was almost even, giving a total of 8 females and 7 males. They had been in the United States for a range of 1 month to 1 year. Six of them had arrived at the beginning of the semester during which

the research was performed, while 4 had already completed another full semester of study in the IEP, and 2 had completed 2 additional semesters. Most of them were working toward undergraduate degrees in fields as diverse as English literature, graphic design, biochemistry, business, criminal justice, computer science, and food science and human nutrition. Some of them had already completed their undergraduate studies and hoped to obtain MBAs in international business. While most of them were in their early twenties, their ages ranged from roughly 18 to 30. The two intact Reading and Writing classes in which this research was performed were taught by the same teacher/researcher. Both classes met for 2 hours a day, 4 days a week, with the first one lasting for 15 weeks and the second lasting for 10 weeks.

3.2 Design (Study 1: Repeated Measures)

The three-day sequence described in Table 1 was performed three times over the course of three weeks in order to investigate what students would notice in three different writing conditions: 1.) when given explicit error corrections of their writing, 2.) when given native-speaker reformulations, and 3.) when given reformulations and asked to talk out loud about them. As can be seen in Table 1, the basic process was the same among all three conditions, the only difference being what happened on Thursday during the 15-minute “comparison stage.” For three weeks, each participant wrote one story each Tuesday, looked at the corrections or reformulations on Thursday, and revised on Friday. All participants had the same amount of time to write the story, engage in some kind of comparison, and then revise it.

TABLE 1**Three-day sequences of the three experimental conditions**

Condition	Tuesday (30 min)	Thursday (15 min)	Friday (20 min)
Error Correction	Write a 30-minute picture description.	Look at explicit error corrections of the essay.	Revise a clean copy of the original essay.
Reformulation	Write a 30-minute picture description.	Compare the essay to a reformulated version.	Revise a clean copy of the original essay.
Reformulation + Think-Aloud	Write a 30-minute picture description.	Compare the essay to a reformulated version while thinking aloud.	Revise a clean copy of the original essay.

As can be seen in Appendix A, an attempt was made to control for the effects of order of condition and writing topic. The participants were divided into three main groups, such that some students would receive corrections the first week, receive reformulations the second week, and do think-alouds the third, while others would receive reformulations first, then do think-alouds, and then receive corrections, and so on. Within each of these groups, the participants were also given different writing prompts, each of which took the form of a picture narrative in comic-strip form. That way, each student would have a chance to write once on each topic and experience each of the conditions one time, but not in the same order as the other students in the class. An example of one of the picture sequences can be found in Appendix B.

The procedure was as follows: For 30 minutes at the end of class on Tuesday, all of the students were given the pictures that had been assigned to them for that week and instructed to write stories describing the pictures. To ensure that they worked through problems with output on their own, they were not allowed to consult with each other or to use dictionaries. At the end of class, the teacher/researcher collected all of the stories along with the pictures. Each story was typed immediately after class, and the errors were coded according to an error classification system of 40 categories adapted from Polio (1997) (in turn adapted from Kroll, 1990). Some expressions were also marked “awkward” if they were not technically incorrect as far as grammar was concerned, but if a native speaker probably would not have expressed the idea in that way. Two independent raters coded each story and obtained reliability at 83.08%, which was slightly higher than the reliability found in Polio (1997). That is to say, of 3481 errors coded, there were 589 disagreements, with a “disagreement” referring to any time the raters coded the same error differently or when one rater coded something as an error while the other did not. Each of the disagreements was discussed until a consensus was reached, and the agreed-upon coding was included in the data analysis. Accidental oversights of unambiguous errors (such as faulty subject-verb agreement, for example) were not counted as disagreements. The full coding system can be seen in Appendix C.

After the error coding had been completed, reformulations of the original stories were typed on separate sheets of paper to be given to the students in the reformulation and think-aloud groups. For the error correction group, extra copies of the participants’ original stories were made, and explicit corrections were written directly on those sheets in purple-colored ink. On Thursday, the students in the error correction group were given

both an unchanged, typed copy of their original story and a copy of that story with the errors corrected on it. The students in the reformulation group received an unchanged, typed copy of their original story along with a copy of the teacher/researcher's reformulation. They were told that they could write on their papers if they wished, but that they would not be able to look at their notes when they rewrote their stories. Those in the think-aloud group were given free time in class to read novels that they had chosen for a reading log project, and they later met with a researcher outside of class. The instructions given to the students can be found in Appendix D, and examples of a student's coded story, a story with corrections on it, and a reformulation can be found in Appendix E.

The students in the think-aloud condition each week signed up for times to meet with the teacher/researcher on Thursday after class. In order to help them to feel comfortable producing a think-aloud protocol, each student was given the opportunity to practice beforehand with an original and a reformulated version of another piece of writing. This warm-up was not recorded in the hope that that would reduce anxiety. The directions given to the participants during the think-aloud can be found in Appendix F. Comments were made by the teacher/researcher during the think-aloud only in order to give instructions, to encourage the participants to keep talking if they had not spoken for a while, and to remind them to speak out loud if they happened to be writing without speaking. Also, some students reached the end of their stories after approximately 10-12 minutes, so they were notified of the amount of time remaining to them if they wished to continue comparing the two versions. (In class, the students in the reformulation and error correction conditions were also encouraged to use the full 15 minutes.)

Immediately after this comparison stage on Thursday, all of the original versions, reformulations, and corrections were collected. Unlike in Qi and Lapkin's (2001) study, retrospective interviews were not done in the hope that this would keep the researcher as neutral as possible and avoid potentially biasing the participants between the comparison and revision stages – in other words, to investigate what the L2 learners noticed without any outside influences. On Friday, the students were given clean copies of the original stories they had written and asked to revise for 20 minutes. When they had finished, everything was collected and typed again, and all of the errors were coded by the same two independent raters. In the end, there were three stories and revisions from each student, with each student having written once about each picture and once in each condition. Errors were tallied with regard to number and type for each story and revision.

At this point, it should be noted that the method of reformulating used in this thesis was somewhat different from what was done in Qi and Lapkin's study. In order to ensure that the same kinds of changes would be made in all of the conditions, the corrections and reformulations for this thesis were based specifically on the errors that had already been coded, with the purpose of reworking instances of linguistic inaccuracy, ambiguity, and awkwardness. As such, we corrected grammatical errors (e.g., choice of preposition, gerund vs. infinitive, subject-verb agreement, punctuation, verb formation, etc.), tried to improve style and cohesion (e.g., by keeping the verbs of the narrative in the same tense, maintaining parallelism, and making sure that pronoun references were not ambiguous), and introduced some new vocabulary in the form of more sophisticated or accurate synonyms for words that were already in the text. However, we did not add any sentences or significantly change the order of existing sentences. We also tried, to

the extent it was possible, not to impose our own writing style or change the meaning of what each student was trying to express. In the end, the main difference between the reformulations and error corrections was a matter of presentation and not related to the kinds of errors that were corrected.

Once the participants had written their stories and revisions, everything was put into columns format, an example of which can be seen in Appendix G. This was done so that the three stages, along with the transcripts of the think-alouds (if applicable), could be compared directly with each other, side by side, to evaluate changes in accuracy from one version to the next. All of the writing was divided into T-units according to guidelines adapted from Polio, Fleck, and Leder (1998), described in Appendix H. Then each T-unit in the participants' revisions was coded for evidence of noticing, with noticing operationalized as an observable correction or partial change at the level of T-units. This was originally done according to the first coding system found in Appendix I; however, later in the study, finding that some of the distinctions were not necessary for our purposes, the coding schema was simplified by collapsing several of the categories. The revised system can also be seen in Appendix I.

In the revised system, according to which the data were finally analyzed, each T-unit could be coded in one of four ways: at least partially changed (+), completely corrected (0), completely unchanged (-), or not applicable (n/a). We considered the "+" and "0" categories to show evidence of noticing, while the "-" category showed no evidence of noticing. The T-units in the "n/a" category were subtracted from the total number of T-units so that we could compare among the conditions how many T-units showed evidence of noticing out of the number of T-units that had contained errors in the

first place. Since all of the individual errors had already been coded, the interrater reliability with the revised system was very high, at over 99%.

3.3 Results (Study 1: Repeated Measures)

After completing the coding for changes in accuracy, evidence of noticing was tallied for each story-revision set, and percentages were taken in order to compare conditions and times. For each participant, the total number of T-units in which there was evidence of noticing (coded + or 0) was divided by the total number of T-units in which some sort of noticing was possible (i.e., those T-units that had contained errors in the original versions). The results comparing these percentages with regard to condition and time can be seen in Tables 2 and 3 below. These data allow us to make preliminary comparisons of the total percentages across conditions, as well as of the revision improvements made by each participant on an individual basis. For instance, Table 2 indicates that Student A showed evidence of noticing in 93.75% of the revised T-units during the Error Correction condition, while showing evidence of noticing in only 83.33% of the revised T-units in the Reformulation condition. As can be seen in the “total” row of Table 2, the participants in the Error Correction condition (in general) showed evidence of noticing on 96.35% of all T-units that originally contained errors. This percentage is higher than the 89.95% of T-units that indicated noticing in the Reformulation condition and the 81.39% in the Think-Aloud condition, suggesting that of the three conditions, error corrections were the most effective in promoting changes in accuracy at the level of T-units, followed by reformulations, and finally think-alouds.

TABLE 2
Comparison of conditions with regard to evidence of noticing (in percentage form)

Participant	Condition		
	Error Correction	Reformulation	Think-Aloud
A	93.75	83.33	92.86
B	100.00	100.00	92.31
C	100.00	93.75	78.26
D	100.00	100.00	81.82
E	100.00	91.67	100.00
F	93.33	93.33	70.59
G	94.44	85.00	91.67
H	100.00	95.24	100.00
I	94.44	88.89	82.61
J	69.23	53.85	64.29
K	100.00	92.86	66.67
L	100.00	100.00	78.95
M	100.00	86.67	71.43
N	100.00	100.00	92.86
O	100.00	84.62	56.52
total n=15	96.35	89.95	81.39

In order to ensure that the counterbalancing attempt to control for order of condition was successful, we also looked at the percentages of corrected T-units for Times 1, 2, and 3. These data are shown in Table 3. Although the “total” row shows a slight increase in percentages over time, an inspection of individual students’ percentages shows that this may be misleading; in fact, a Friedman Test on the ranked percentages did not find significant differences according to time, suggesting that the counterbalancing attempt was successful and the results regarding comparison of conditions were not affected by improvement over time. The results of this test can be seen in Table 4.

TABLE 3
Comparison of times with regard to evidence of noticing (in percentage form)

Participant	Time		
	Time 1	Time 2	Time 3
A	93.75	83.33	92.86
B	100.00	92.31	100.00
C	78.26	100.00	93.75
D	81.82	100.00	100.00
E	91.67	100.00	100.00
F	93.33	93.33	70.59
G	94.44	85.00	91.67
H	95.24	100.00	100.00
I	88.89	82.61	94.44
J	69.23	53.85	64.29
K	92.86	66.67	100.00
L	78.95	100.00	100.00
M	71.43	100.00	86.67
N	92.86	100.00	100.00
O	100.00	84.62	56.52
total n=15	88.18	89.45	90.05

TABLE 4
Comparison of times with regard to evidence of noticing
Friedman Test of ranked percentages

mean rank of each time			test statistics			
Time 1	Time 2	Time 3	N	chi-square	df	asympt. sig.
1.87	1.93	2.20	15	1.057	2	.590 (n.s.)

The three conditions were also compared with respect to evidence of noticing (or revision accuracy) by performing a Friedman Test, ranking the percentages of revised T-units that contained at least one correction or change. These results are presented in Table 5, and they indicate that the Error Correction condition, with a mean rank of 2.77, had the most accurate revisions, while the Think-Aloud condition had the least, with a mean rank of 1.40. The results were significant overall (asyp. sig. 0.000), which would allow the rejection of a null hypothesis that projected no differences according to condition. The strength of association, according to the η^2 formula for Friedman tests from Hatch and Lazaraton (1991), was 0.3765, indicating that approximately 38% of the variability could be accounted for by the three different conditions.

TABLE 5
Comparison of conditions with regard to evidence of noticing
Friedman Test of ranked percentages

mean rank of each condition			test statistics			
EC	R	TA	N	chi-square	df	asyp. sig.
2.77	1.83	1.40	15	16.566	2	.000*

After finding statistical significance with the Friedman Test, Wilcoxon Signed Ranks Tests were performed in order to investigate the degree and direction of differences between pairs of conditions. These results can be seen in Tables 6, 7, and 8. Table 6 shows a significant difference between the Error Correction and Reformulation

conditions. The percentages of T-units with evidence of noticing in the Error Correction condition outranked or tied those in the Reformulation condition, and η^2 (the strength of association) was a strong .5620. Table 7 shows that the Reformulation condition was significantly better than the Think-Aloud condition, but with a smaller strength of association (.2823) and not as much statistical significance. Between the Think-Aloud and Error Correction conditions, Table 8 shows a significant difference and a high strength of association at .7223. Mirroring the differences in mean ranks shown in Table 5, there appears to be a greater difference between Error Correction and Reformulation than between Reformulation and Think-Aloud.

TABLE 6
Comparison of the Error Correction and Reformulation conditions
Wilcoxon Signed Ranks Test

	<u>N</u>	<u>mean rank</u>	<u>sum of ranks</u>	<u>Z</u>	<u>asyp. sig.</u> <u>(2-tailed)</u>	<u>η^2</u>
Negative Ranks	0 ^a	.00	.00			
Positive Ranks	10 ^b	5.50	55.00			
Ties	5 ^c					
Total	15					
				-2.805 ^d	.005*	.5620

a. Error Correction < Reformulation

b. Error Correction > Reformulation

c. Reformulation = Error Correction

d. Based on negative ranks.

TABLE 7

**Comparison of the Think-Aloud and Reformulation conditions
Wilcoxon Signed Ranks Test**

	<u>N</u>	<u>mean rank</u>	<u>sum of ranks</u>	<u>Z</u>	<u>asympt. sig. (2-tailed)</u>	<u>η^2</u>
Negative Ranks	10 ^a	9.50	95.00			
Positive Ranks	5 ^b	5.00	25.00			
Ties	0 ^c					
Total	15					
				-1.988 ^d	.047*	.2823

a. Think-Aloud < Reformulation

b. Think-Aloud > Reformulation

c. Reformulation = Think-Aloud

d. Based on positive ranks.

TABLE 8

**Comparison of the Think-Aloud and Error Correction conditions
Wilcoxon Signed Ranks Test**

	<u>N</u>	<u>mean rank</u>	<u>sum of ranks</u>	<u>Z</u>	<u>asympt. sig. (2-tailed)</u>	<u>η^2</u>
Negative Ranks	13 ^a	7.00	91.00			
Positive Ranks	0 ^b	.00	.00			
Ties	2 ^c					
Total	15					
				-3.180 ^d	.001*	.7223

a. Think-Aloud < Error Correction

b. Think-Aloud > Error Correction

c. Error Correction = Think-Aloud

d. Based on positive ranks.

Incidentally, similar results were obtained upon considering the percentages of revised T-units in which all of the errors were corrected or changed (corresponding only to a "0" coding). The Error Correction condition had the greatest percentage of

completely corrected T-units, while the Think-Aloud condition had the smallest.

According to the “total” row of Table 9, participants completely corrected 47.02% of all the T-units that originally contained errors in the Error Correction condition, 31.88% in the Reformulation condition, and 22.22% in the Think-Aloud condition. Again, a Friedman Test of ranked percentages (Table 10) showed the Error Correction condition to be better than Reformulation, and Reformulation to be better than Think-Aloud. The results were significant, and as can be seen in Table 11, there does not appear to be an effect of time (order of condition).

TABLE 9
Comparison of conditions with regard to complete correction (in percentage form)

Participant	Condition		
	Error Correction	Reformulation	Think-Aloud
A	43.75	16.67	14.29
B	25.00	20.00	23.08
C	14.29	12.50	8.70
D	90.00	83.33	45.45
E	41.67	25.00	40.00
F	53.33	66.67	29.41
G	22.22	10.00	25.00
H	69.23	52.38	10.00
I	11.11	27.78	17.39
J	30.77	15.38	7.14
K	73.33	35.71	44.44
L	83.33	30.00	26.32
M	43.75	20.00	4.76
N	81.25	47.37	28.57
O	22.22	15.38	8.70
total n=15	47.02	31.88	22.22

TABLE 10
Comparison of conditions with regard to complete correction
Friedman Test of ranked percentages

mean rank of each condition			test statistics			
EC	R	TA	N	chi-square	df	asyp. sig.
2.73	1.87	1.40	15	13.733	2	.001*

TABLE 11
Comparison of times with regard to complete correction (in percentage form)

Participant	Time		
	Time 1	Time 2	Time 3
A	43.75	16.67	14.29
B	20.00	23.08	25.00
C	8.70	14.29	12.50
D	45.45	90.00	83.33
E	25.00	40.00	41.67
F	53.33	66.67	29.41
G	22.22	10.00	25.00
H	52.38	10.00	69.23
I	27.78	17.39	11.11
J	30.77	15.38	7.14
K	35.71	44.44	73.33
L	26.32	83.33	30.00
M	4.76	43.75	20.00
N	28.57	81.25	47.37
O	22.22	15.38	8.70
total n=15	29.80	38.11	33.21

3.4 Analysis of Think-Alouds

At the beginning of this study, it was hypothesized that reformulations and think-alouds might lead to more accurate revisions by encouraging more active search and deeper processing of corrections. Accordingly, the main part of this thesis compares three conditions (Error Correction, Reformulation, and Think-Aloud) with respect to the percentages of revised T-units that show evidence of noticing. However, the additional think-aloud data generated from this inquiry offer many other avenues for data analysis. Therefore, after the completion of these first analyses, an investigation of quality of noticing was begun in order to find out how instances of noticing of different qualities might be related to changes made in revisions. In order to do this, the analysis was restricted to the think-aloud data and “noticing” was operationalized in a new way. This time, noticing was not operationalized as a change in accuracy made in a revision at the T-unit level. Instead, since there was access to what the participants had said about each error in the think-alouds, noticing was operationalized as a verbalization related to an error, and a three-tiered coding system was used to classify each error from the original story.¹

On the first tier of this new system, each original error was coded either +N or – N (for whether or not it was noticed in the think-aloud), and either +C or –C (for whether or not it was corrected), or H (for when something was changed but not completely

¹ Of course, given that people can notice things without necessarily speaking about them, we cannot assert that the participants’ verbalizations provide evidence of everything they noticed. Likewise, we cannot necessarily assert that a more elaborate verbalization corresponds to deeper processing (or higher quality of noticing) since people can look at things without speaking and reflect on them deeply without verbalizing their thoughts. However, it did end up being the case that what we labelled as “higher quality” noticing was associated with corrections more often than not. Presumably, we could have gotten important additional information through videotaping or tracking eye movements, and these would certainly be interesting avenues for future research.

corrected). This tier of coding had an interrater reliability of about 99%. Then, on the second tier, quality of noticing was assessed by looking back at all of the noticed (+N) errors and classifying what kind of comment was made about each in the think-aloud. This coding system had 85.24% interrater reliability. The categories were as follows, and an example of each can be found in Appendix J.

M: mentioning an error or correction without a reason or rereading with special emphasis

SP: making note of a misspelling

ML: using metalanguage without a reason

SM: recognizing a “stupid mistake”

R: providing a reason for a correction

LN: making note of a new lexical item

LO: making note of a familiar lexical item

NR: not being able to provide a reason

RJ: rejecting a change

WR: providing an invalid reason for a correction

RD: simply reading a correction aloud without comment

Finally, keeping in mind the output hypothesis and the idea of “noticing the gap,” the third tier of the coding system was included to try to make note of times when participants were aware of their own initial output problems *and* aware of the differences between IL and TL when they saw the reformulations. In Qi and Lapkin’s study, it

seemed to the researchers as though the participants experienced a “sense of lack of fulfillment” when they could not solve language problems while writing, and in fact, most of the problems that they talked about during the writing stage were then noticed during the comparison stage (p. 289). The participants often made exclamations when they realized that there were differences, accepting the reformulations as better and mentioning that they had wanted to express their ideas in a better way, but had not known how. In view of this, we also assumed that a learner’s inability to come up with the language needed to express an idea would push him/her to be on the lookout for relevant input in the future. Since our participants did not think aloud during the initial writing task, it was not possible for us to check whether the problems they noticed at that time were noticed more often than not during the comparison stage. However, it was possible to examine the verbalizations made during the comparison stage and make note of times when the participants indicated they had “noticed the gap,” mentioning differences between what they had originally wanted to produce, what they were actually able to produce, and what the native speaker produced. Thus, if a participant said something along the lines of, “Oh, I wanted to use that word, but I couldn’t remember it!” we marked this as evidence of noticing the gap and hypothesized that the learner might be more predisposed to remember the correction and use it later in the revision.

3.4.1 Association between noticing² (and quality of noticing) and correction

There was a clear association between noticing (verbalizations made) during the think-alouds and corrections made on the revisions. Table 12 shows that if an error was noticed, then it was more likely to be corrected than not. Inversely, if something was not noticed, then it was less likely to be corrected. The converse is also true: If an error was corrected, then it was more likely to have been noticed than not.

In a preliminary attempt to explore the relationship between quality of noticing and correction, two categories were chosen as partially representative of “high quality” noticing, with the assumption that they would be a subset of the “substantive” kind of noticing from Qi and Lapkin: ML (the use of metalanguage) and RE (provision of a reason). Looking at the quality of noticing in the last two columns of Table 12, it appears that providing a reason for a correction or using metalanguage about it during the think-aloud was associated more with making a correction in the revision than with not making one. All of these results seem to confirm Qi and Lapkin’s findings.

3.4.2 Quality of noticing and noticing the gap

For reasons which will be discussed below, the coding from Tiers II and III has not been analyzed in depth to compare different qualities of noticing with each other (Tier II) or to investigate how noticing the gap (Tier III) may be related to subsequent correction in the revisions. In Qi and Lapkin’s study, a simple distinction was made between noticing with a reason (called “substantive” noticing) and noticing without a

² It is important to keep in mind that “noticing” here refers *not* to changes made in revisions, as it was operationalized for the comparison of verbalizing and non-verbalizing conditions, but rather to statements made during the think-alouds of the comparison stage.

reason (“perfunctory” noticing). Using a larger number of distinctions in Tier II, the categories that were used in this thesis were not always so easy to divide in that way. Problematic aspects of this coding system will be considered in Section 5.2: Implications for research methodology.

TABLE 12
Associations in the think-aloud data between noticing and correction and between “high quality” noticing and correction

Participant	relationship between noticing (N) and correction (C)				relationship between high quality noticing (ML/RE) and correction	
	+N+C	+N-C	-N+C	-N-C	ML/RE +C	ML/RE -C
A	22	7	0	1	1	0
B	19	6	5	15	3	2
C	24	10	0	7	2	1
D	12	3	0	5	3	0
E	21	5	0	1	6	0
F	20	0	2	9	9	0
G	13	15	1	4	1	0
H	17	3	4	5	3	1
I	12	10	2	10	4	4
J	13	2	26	22	4	0
K	21	22	5	5	10	11
L	18	2	12	6	7	0
M	18	1	10	24	9	1
N	14	0	9	20	6	0
O	17	2	2	3	10	2
total n=15	261	88	78	137	78	22

+N = noticed
+C = corrected

- N = not noticed
- C = not corrected

ML/RE = use of metalanguage or
provision of a reason (high quality)

3.5 Problems Leading to Study 2 and Rationale for Modifications in Design

Originally, it was hypothesized that the Reformulation and Think-Aloud conditions would be more effective than Error Correction in promoting noticing because of additional search and verbalization components. The opposite turned out to be the case in the first study, and it seemed plausible that the results might have had to do with excessive cognitive load. Given that the participants in the Error Correction condition did not have to search for their corrections and could therefore spend more time and devote more cognitive resources to remembering differences, it seemed possible that they might have been able to use memorization strategies. Furthermore, since they rewrote their stories on the day immediately following the comparison stage, they might have been able to remember the corrections easily regardless of whether or not they had actually understood them. In order to either confirm or deny these speculations, post-study debriefings were conducted with six of the participants soon after the first study was completed, with the following seven questions:

- 1.) Which activity was the easiest for you to do?
- 2.) Which was the most difficult?
- 3.) Which made it easiest to remember the corrections?
- 4.) Did you use any strategies when you were comparing?
- 5.) Do you think your strategies changed over time?
- 6.) Which activity did you like best?
- 7.) Which one do you think was the most useful?

In the post-study interviews, some of the students did, in fact, mention having tried to memorize the changes they had seen in the Error Correction condition. Interestingly, some had tried to remember not merely what the corrections were, but even what the writing on the paper had looked like and where the errors had been located on the page. One participant said that the written error corrections were more “impressive” and easier to remember visually. Another said that she had counted the errors in the Error Correction condition and tried to remember how many there were in each line of her story, while someone else noted that she had wanted to take the time to memorize the corrections during the Think-Aloud condition, but had not been able to do so because she had had to concentrate her efforts on talking. Another participant mentioned having had enough time in the Error Correction condition to read through the clean copy he had been given, try to make the corrections himself, and then go back and check them. Appendix K presents some more illuminating statements made by the participants during the post-study interviews, indicating not only that memorization strategies were attempted, but also that the requirement to talk aloud in a second language might have divided the participants’ cognitive resources while they were trying to complete the comparison task.

With these issues in mind, some design modifications were made for a second study in an attempt to temper the participants’ ability and inclination to make use of memorization strategies. First of all, the repeated-measures design was abandoned, and data were collected from a greater number of participants. In view of the fact that the participants would complete the three-day sequence only one time each, it seemed unlikely that they would have as much of a chance to recognize the usefulness of memorization strategies. In other words, even though they would be told that they had to

revise their stories, they would not know firsthand exactly what this was like until they did so. Another important change for the purpose of reducing the use of memorization strategies was the inclusion of more time in between the comparison and revision stages. The second study still involved a three-day sequence, but instead of using a Tuesday-Thursday-Friday sequence in which everything was completed during the same week, it was done on Monday-Wednesday-Monday or Tuesday-Thursday-Tuesday. Finally, in order to establish how well the participants were able to revise their stories on their own, a true Control condition (X) was added. Those in the Control condition completed exactly the same activity as those in the other three conditions, except that during the comparison stage they looked at their uncorrected stories for 15 minutes by themselves while the other participants were looking at corrections or reformulations.

Chapter 4

STUDY 2 (NON-REPEATED MEASURES DESIGN)

4.1 Participants (Study 2: Non-Repeated Measures)

The participants in the second study were 54 ESL students from a variety of levels. Most of them came from the IEP (Intensive English Program) and EAP (English for Academic Purposes) programs at the same large Midwestern university, while an additional 10 participants came from an ESL class at a local community college. Of the university participants, 23 came from the IEP, with 16 from Level 300 (high intermediate) and 7 from Level 400 (advanced), and 21 came from the EAP, with 17 from Level 093 (Academic English Grammar and Composition for Non-Native Speakers) and 4 from Level 095 (Academic English Composition for Non-Native Speakers). Native languages included mostly Korean and Japanese, but also Chinese, Portuguese, Spanish, and French. None of the students had participated in the first study, and none of the classes in which the research was conducted were taught by the researcher. The participants were randomly divided into conditions within each class.

4.2 Results (Study 2: Non-Repeated Measures)

After all of the participants had completed the three-day writing sequence, changes in accuracy were coded and evidence of noticing was tabulated for each story-revision set. Then percentages were calculated in order to compare the four conditions

(Error Correction, Reformulation, Think-Aloud, and Control) with regard to evidence of noticing shown in the revisions. Again, the total number of revised T-units in which there was evidence of noticing (coded + or 0) was divided by the total number of T-units in which some sort of noticing was possible (i.e., those T-units that had contained errors in the original versions). The results can be seen in Table 13.

On a preliminary straight comparison of percentages, even with the design modifications that had been made, the Error Correction condition still seemed to enjoy the most accurate revisions, with 87.55% of the T-units showing evidence of noticing. As expected, the participants who had received no feedback on their writing in the Control condition wrote the least accurate revisions, showing evidence of noticing errors in only 55.16% of their revised T-units. Since a normal distribution was not assumed and because of large differences between the conditions with regard to standard deviations, it would not have been useful to calculate effect sizes for these data. However, it is interesting to note that the Reformulation and Think-Aloud conditions' results look more similar to each other in this study than in the first, and in fact, they seem to have switched places in the order on a comparison of straight percentages, with the participants in the Think-Aloud condition seeming slightly to have outperformed those in the Reformulation condition.

On the other hand, if we compare the conditions by *ranking* the percentages of T-units showing evidence of noticing for each story-revision set, the order of conditions echoes that of the first study. Using Condition as the grouping variable in a Kruskal-Wallis Test, the mean rank of percentages in the Error Correction condition comes out on

top, followed by Reformulation, Think-Aloud, and finally Control. The results are significant overall and can be seen in Table 14.

TABLE 13

Comparison of conditions with regard to evidence of noticing (in percentage form)

Error Correction	Think-Aloud	Reformulation	Control
87.55	72.94	70.51	55.16

TABLE 14

Kruskal-Wallis nonparametric test

	condition				total	test statistics		
	EC	R	TA	C		chi-square	df	asympt. sig.
mean rank	42.63	28.14	26.84	15.63				
N	12	11	16	15	54			
percent of T-units noticed						19.676	3	.000

Applying Mann-Whitney tests to check for two-tailed significance in the differences between the conditions in order, the difference in mean rank between Error Correction and Reformulation was significant at .025, and the difference between Think-Aloud and Control was significant at .013. However, the difference in mean rank between the Reformulation and Think-Aloud conditions was not significant.

Incidentally, it may also be interesting to note that each of the conditions in the second study's modified non-repeated measures design had less accurate revisions than

its corresponding condition from the first study's repeated measures design, when less time intervened between the comparison and revision stages. While overall about 96% of T-units showed evidence of noticing in the Error Correction condition in Study 1, only about 88% in the Error Correction condition showed such evidence in Study 2. For the Reformulation condition, the correspondence was 90% to 71%, and for the Think-Aloud condition it was 81% to 73%. This might seem to suggest that memory was a factor, but it is important not to jump to conclusions since the participants were not exactly the same in the two studies, the second study including a wider range of L2 proficiency.

Chapter 5

DISCUSSION

5.1 Discussion of Research Questions

The first parts of the two studies in this thesis involved a comparison of different writing conditions with respect to the noticing and subsequent revision accuracy they were able to promote. The second part of the first study consisted of an investigation of what L2 learners notice and how that noticing is related to changes made in revisions. Thus, to keep the order of presentation constant, it makes sense to discuss research questions 3 and 4 first, followed by research questions 1 and 2.

5.1.1 Research question 3: Do students notice more when comparing their essays to reformulated versions as opposed to versions with explicit error corrections?

Despite logistical problems associated with using reformulations, we were interested in Qi and Lapkin's suggestion that reformulations' ability to serve as a relevant model of native-like writing might be a helpful pedagogical tool and a better alternative to less-than-optimal written error corrections. We assumed, along with Qi and Lapkin, that corrective feedback might work better when learners can not only pay attention to form, but also make comparisons between their IL and a TL model. Reformulation seems to be in accordance with ideas about the importance of positive and negative evidence and a focus on both meaning and form. It also seems to induce both error

analysis and cognitive comparison, and as such, we thought it might lead to a more analytical orientation, more metalinguistic awareness, and a greater development of cognitive strategies for noticing. Therefore, in response to the third research question, we hypothesized that the active search and cognitive comparison involved in finding the differences between two intact versions of writing (in the reformulation condition) would induce more noticing and greater linguistic accuracy in revisions than would occur with explicit error correction.

Surprisingly, the results indicate exactly the opposite of what we expected. Based on the data, reformulations are *not* more helpful than explicit error corrections for the purpose of producing revisions with greater accuracy and evidence of noticing. The fact that the participants in the Reformulation condition outperformed those in the Control group suggests that reformulations are helpful. However, the participants in the Error Correction condition consistently produced the most accurate revisions (with the most evidence of noticing) at the level of T-units.

When interpreting these results, there are several factors to keep in mind: namely, the perceptual salience of the written error corrections, the amount of work that had to be done in each condition (and the corresponding allocation or division of cognitive resources), time limitations, the amount of time between stages (related to possible memory concerns), and any potential long-term effects.

First of all, it is important to note that the explicit corrections, which were written on the students' papers in purple ink, actually made the differences more perceptually salient. Since participants in the Error Correction condition did not have to worry about searching for differences or talking about what they were doing, perhaps they were able

to devote their cognitive resources to understanding and remembering the corrections. The active search component of the Reformulation and Think-Aloud conditions may have caused participants' cognitive resources to be divided. They clearly had more work to do, and in the post-study interviews, the participants' comments corroborated the idea that finding the differences in the reformulations might have been more difficult.

It is also unclear how much of an effect time limitations had on the performance of a task in conditions that might have required different amounts of time. All of the participants were given 15 minutes to complete the comparison stage regardless of whether they simply had to look at corrections or search for them. However, it presumably takes more time to search for differences and think about them (R, TA) than simply to look at differences that have already been clearly identified (EC). Unfortunately, simply giving participants as much time as they needed would not have solved the problem, either. If it biased results in the other direction, it might be possible for one to argue that more time on task was the deciding factor.

In addition, even though the amount of time between stages was increased in order to reduce possible memorization effects in the Error Correction condition, it is possible that it was not increased enough. The possibility of using memorization strategies more effectively in some conditions than others might still have played a role in the results, not to mention the possibility of unintentional memories (e.g., visual memory of perceptually salient features) having an effect.

Finally, the results do not reveal anything about the long-term effects of noticing in each condition. It is not possible to say whether or not the search involved in the reformulation and think-aloud conditions led to deeper processing, more metalinguistic

awareness, or the development of cognitive strategies for noticing. Presumably, this might happen over a longer period of time and with repeated practice. Qi and Lapkin pointed out that the learners in their study occasionally noticed corrections in the reformulations and gave appropriate reasons for them without subsequently incorporating them into their revisions. They suggested that even though this experience of noticing and understanding did not help the participants immediately in their revisions, it might have helped them to notice relevant features in future input or output. This may also be true in the case of our participants, who occasionally did not incorporate corrections into their revisions even though they had shown themselves to understand them in their verbalizations.

5.1.2 Research question 4: Does the use of think-aloud protocols affect the number of linguistic features that students notice and that subsequently make their way into the final version of the written text?

We hypothesized that noticing might be positively affected by thinking aloud since the requirement to verbalize might encourage participants to engage in additional reflection and problem solving in order to figure out the reasons behind the differences they found. Evidence that the revisions in the Think-Aloud condition had improved more in accuracy than those in the other conditions could have been used to support this position, but no such evidence was found. Apparently, thinking aloud while comparing an original story to a reformulated version of it *reduces* the number of T-units in which

errors are corrected. However, as with the previous research question, there are additional issues to keep in mind.

As before, time limitations may have been a factor, but there are also special considerations related specifically to the task of thinking aloud. For example, there could have been reactivity and nonveridicality in the form of an inappropriate level of verbalization (associated with social communication or description of activities), and the use of an L2 automatically introduced speaking proficiency as an influential factor. Moreover, looking back on our coding system for identifying noticing, it is clear from what the participants said during the think-alouds (and how that was related – or not related – to what appeared in the revisions) that we should take into account both our inability to detect all noticing and our inability to know for certain whether or not a change in accuracy constituted noticing. This is true not only for the Think-Aloud condition, but for the other conditions as well.

Since think-alouds are known to increase the amount of time it takes to complete a task, it is possible that the participants in the Think-Aloud condition were not able to devote any time to trying to remember the corrections, even if they understood them well at the time of comparison. In Qi and Lapkin's study, they noted that just because the participants accepted reformulations for the right reasons did not mean that they would remember them in the revision stage. Making reference to Robinson (1995), they stated, "Even noticing with comprehension may need some reinforced rehearsal in memory" (p. 295). This reinforced rehearsal may have been what our Error Correction condition provided. In a discussion of the potential benefits of thinking aloud, Ericsson and Simon (1993) brought up the question of whether such benefits could "offset any disadvantage

from the additional time taken to verbalize the information” (p. xxxi). In the case of our study, the answer is apparently not, and this serves to underscore the importance of not overlooking the issue of time in L2 research methodology.

Also important in L2 research methodology employing think-alouds are the issues of social communication and what Ericsson and Simon have called an inappropriate level of verbalization. In this particular study, by the time a participant produced a think-aloud protocol, the researcher (a native speaker “expert”) had already read, analyzed in detail, and corrected his or her work. The researcher was in the room at the same time, and the participant knew not only that she was listening to his or her assessments of the corrections she had made, but also that she had an automatic understanding of the information that the participant wanted to learn. Thus, even though the researcher sat apart from the participants, and even though the instructions informed the participants that the researcher would not answer any questions or talk with them, it still seemed as though they were sometimes engaging in social interaction instead of simply speaking their thoughts out loud. This, and the way they constructed the task, may have encouraged them to explain things that they would not have explained otherwise, and it may have forced them to think explicitly about processes that were normally automatic for them. They were not simply focused on the task; they were also concentrating on saying coherent things that could be understood by another person. Ericsson and Simon have warned that this may not represent true online thinking and may disrupt underlying thought processes.

The fact that the participants had to produce their think-aloud protocols in an L2

must have increased their cognitive load even more. One of the reactivity-causing factors mentioned by Stratman and Hamp-Lyons (1994) is the limited short-term memory (STM) capacity for talking and attending at the same time. Researchers often assume, along with Ericsson and Simon, that although time may be affected, underlying thought processes themselves should not be affected by the necessity of verbalization as long as a verbal code in which they can be expressed is readily available. However, in an L2, coming up with the terms and grammar necessary for expressing thoughts is not a highly automated process; it requires considerably more effort than it does in an L1 and may affect the cognitive processes involved. Depending on proficiency, the use of an L2 may also affect the kinds of thoughts a learner is able to express.

Especially for L2 learners, producing a think-aloud protocol may thus be equivalent to carrying out an additional task and may affect their ability to concentrate on performing the primary task efficiently. From the point of view of a participant in our study, the minimal task may not have been simply to find differences (and, incidentally, also to speak thoughts out loud), but rather to *have things to say* about finding differences and to *concentrate on finding the language* to express their ideas. Given this extra task and the correspondingly heavier burden on STM, L2 verbalization might have competed with and interrupted the primary task, leading to a loss of information from STM.

According to Russo, Johnson, and Stephens (1989), “prolonged attention to items in STM to allow verbalization will be disruptive of tasks that impose high loads on STM” (p. 759). Providing some support for this are the words of one of the participants from the post-study debriefings:

Uh, it is, uh, when I speak English, I am very worry about, worry about grammar, so even though I find out my mistake, I, I have to, actually, it is, my mistake is not important in think-aloud because I concen- How can I explain to you? So... uh, even though I found a, my mistake, I... yeah, it is hard to memorize my mistake.... I don't need to speak something, so I can memorize easily, but think-aloud is, uh, I, uh, notice my mistake, and then I have to do, tell you, and then I forgot.

5.1.3 Research question 1: What do L2 learners notice as they compare their text to a reformulated version while thinking aloud?

The verbalizations in the think-aloud protocols made it clear that L2 learners are successful in noticing a wide range of error types. However, pinning down the phenomenon of “noticing” may be problematic from a methodological standpoint. In our study, wanting to find out what the participants were noticing in all of the conditions (and not just when they talked aloud), we used changes in accuracy from story to revision as a way of ascertaining what might have occurred during the comparison stage. Then we used changes in accuracy coding (+/0/-/na) to compare the conditions with respect to “evidence of noticing.” We were also able to look at changes in the quantities of various kinds of errors from the stories to the revisions by looking at the error tallies we compiled for each participant. An example of an error tally sheet can be found in Appendix E.

Not too surprisingly, though, when the extra verbalization data from the think-alouds provided access to at least some of what the participants were noticing, we found that the changes in accuracy and the verbalized instances of noticing did not always correspond. In other words, just because participants talked about something – even if they talked about it in depth and displayed understanding – that did not mean that they would remember and change it in the revision. Conversely, just because they changed something in the revision did not mean that it was related to something they had said in the think-aloud. Our coding system was necessarily imprecise and had limitations. We often marked “-” (no evidence of noticing) when the think-alouds clearly demonstrated that the participants had noticed errors, and we sometimes marked “+” (evidence of noticing) when no mention was made of the errors in the think-alouds. Participants often revised what they apparently had not noticed and did not revise what they apparently had noticed.

What this means, of course, is that we do not know precisely what the participants noticed in the Reformulation and Error Correction conditions. We can assume, by extension, that not everything that was noticed showed up in the revisions and that some of the linguistic items that did show up in the revisions were unrelated to what the participants had seen during the comparison stage. Essentially, we need to make two provisos: 1.) that we were unable to detect or infer all noticing, and 2.) that in the absence of other evidence, we could not know definitively whether a revision change was related to noticing in the comparison stage. A direct (or nearly direct) correspondence between a reformulation and revision sometimes made it clear that noticing had influenced the revision (e.g., if someone wrote “lish,” trying to use the new word “leash” that he or she

had seen in the reformulation). However, other times it was much more difficult to tell (e.g., if someone changed “on” to “in” after (possibly) seeing “at” in the reformulation). Another observation that may be worthy of note is that several participants in the Think-Aloud condition wrote corrections that they had presumably seen in the reformulations without mentioning them out loud.

5.1.4 Research question 2: How is noticing related to revision changes completed after comparing the original and reformulated versions of a story?

One of the original interests when designing the studies for this thesis was to see if it would be possible to confirm quantitatively Qi and Lapkin’s assertion that the quality of noticing experienced while comparing an original story to a reformulation could have direct implications for the revision of that story. According to Qi and Lapkin, noticing with a reason might have more of an impact on learning than noticing without understanding. It should be pointed out that we do not intend to suggest a cause-effect relationship between noticing and subsequent linguistic accuracy based on the results of our study, as Qi and Lapkin may have implied in theirs. It is not possible for us to declare that the fact that participants verbalized (noticed) things – or the way in which they did so – actually caused them to be changed in the revisions. This may have had to do with other factors, including the learners’ developmental readiness to notice and/or acquire forms and their ability to talk about them. It is interesting, though, that our data suggest associations. For instance, errors that have been noticed are more likely to be corrected. It also seems to be the case that if L2 learners use metalanguage or give a

reason for an error, it is more likely to be corrected than not. (In other words, what we labelled as “high quality” noticing was associated with corrections more often than not.)

It should also be noted that the fact that the participants in the Error Correction condition wrote more accurate revisions than those in the Think-Aloud condition does not invalidate the theory that higher quality noticing is related to more corrections and uptake. What it tells us is that, in the short term at least, and given a very fixed amount of time on a three-stage writing task, students produce more accurate revisions with error corrections than with reformulations. When all is said and done, the participants in the Error Correction condition enjoyed corrections that were more perceptually salient, a lighter workload, and a correspondingly lesser division of cognitive resources. These factors may have let them exploit “detection plus rehearsal in short term memory” (a definition of noticing from Robinson, 1995) to a greater extent than the participants in the other conditions were able to. Altogether, regardless of who achieved better quality noticing in the long run, the circumstances of the EC condition might have ended up outweighing the importance of whatever search, evaluation, and cognitive comparison the reformulations might have encouraged participants to do. As an alternative explanation, the circumstances might have allowed the EC participants actually to engage in *more* evaluation and cognitive comparison since they did not have to spend their time searching for differences.

5.2 Implications for Research Methodology

Through the discussion of these research questions, we have already seen that it is essential to consider at least three factors in L2 research methodology: 1.) the effects of time (and the apparently impossibility of controlling for it completely), 2.) the general limitations and benefits of different levels of verbalization, along with the special constraints that may accompany verbalization in an L2, and 3.) the difficulty of pinning down the phenomenon of noticing. In addition, when attempting to get at “quality of noticing” as a construct, it is important for researchers to realize, first of all, that they cannot observe everything that is happening inside learners’ heads, and relatedly, that the distinctions they impose may not be so clear-cut in reality. It is also important to recognize that the *type* of error a learner has noticed may have its own effect on how much is verbalized about it; “substantive” noticing may not always be necessary or even possible for certain kinds of errors. Finally, L2 writing researchers should be aware of the problems involved in comparing straight numbers or percentages of errors noticed or changed in revisions.

5.2.1 *Problems with attempts to distinguish between noticing of different qualities*

As mentioned above³, Qi and Lapkin made a simple distinction between “substantive” noticing (with a reason) and “perfunctory” noticing (without a reason). However, we found that the categories we used in Tier II of our coding system were not always so easy to divide in that way. For instance, recognizing a “stupid mistake” might be somewhat substantive since a learner might know a reason on some level even without

³ 3.1.4.2 *Quality of noticing and noticing the gap*

stating it explicitly (e.g., “Right! They were worried! (laughs) Why I put ‘worry’? Yeah, right! Worried.”). In this case, the participant never said why “worried” was better than “worry,” but the noticing was not merely perfunctory or glossed over without any thought or understanding. Whether this meant that the participant did not need to think actively about anything additional in order to figure out the mistake or whether it meant that he or she simply was not verbalizing all of his or her thoughts, the result would be that (in appearance to the researcher, as far as the verbalization data were concerned) the participant seemed to understand the mistake somewhat automatically. Likewise, in noticing a spelling mistake, a learner would simply have to note the misspelling, and a further (more substantive) explanation of a reason would be unnecessary.

A clear-cut case of perfunctory noticing would seem to be merely reading a correction without commenting on it at all (RD). Mentioning a correction and repeating it with emphasis but without saying anything additional (M) might also seem to be perfunctory. However, here it becomes difficult to draw the line. What is the difference between simply mentioning a mistake (“Oh, *look at.*” or “Oh, I need *at.*” - M) and using metalanguage without a reason (“Oh, I need a preposition.” - ML)? Is the first perfunctory and the second substantive just because a label is used? Are they both perfunctory or both substantive? One problem seems to be that there is too much gray area in between the clear-cut, extreme cases of perfunctory and substantive noticing. Effort seems to be a factor as well. In our coding system, we classified a verbalization as exemplifying “lack of reason” only if a participant actually said something along the lines

of, “I don’t know.” But even this could possibly be substantive if the participant employed considerable mental resources before eventually giving up.

Ultimately, this becomes a question of which phenomena we are trying to identify when we classify noticing as perfunctory or substantive. As far as a researcher’s coding is concerned, “substance” might have to do with the quantity or completeness of verbalization. That may in fact be correlated with depth of processing and quality of noticing, and it may end up being correlated with more corrections in revisions. If so, that would be interesting to know for pedagogical purposes. Nonetheless, while an analysis of what and how much is verbalized can certainly provide clues and give a researcher a better idea about quality of noticing, it is not possible to tell definitively how deeply a correction has been processed and why it has been remembered in a revision based only on this. Perhaps levels of noticing could be divided into four areas for future research: 1.) the most substantive, including an explanatory reason and evidence of relatively complete understanding (e.g., RE), 2.) somewhat substantive, with evidence of at least some level of understanding or effort even if a correct or explicit reason is not given (i.e., the gray area discussed above, e.g., SM, ML, M), 3.) purely perfunctory, with no evidence of understanding displayed (e.g., RD), and 4.) no evidence of noticing. Another possibility for describing these levels might be: 1.) the most substantive, including possible evidence of processing depth, along with elaboration, 2.) substantive, including possible evidence of processing depth but no elaboration, 3.) purely perfunctory, with no evidence of processing depth or elaboration, and 4.) no noticing. These distinctions would not completely address the issue, but they might be

improvements since they leave room for the gray area between the extreme cases and incorporate some account of how much effort is made.

5.2.2 Additional effects of error type on the construct “quality of noticing”

It is also important to keep in mind that the type of error (e.g., verbal aspect vs. spelling) can have an effect on how much a participant verbalizes; what kinds of noticing, awareness, and understanding are likely; and whether or not a substantive explanation is possible or necessary. Qi and Lapkin used three categories when considering which errors were correctly or incorrectly revised: lexical, form, and discourse. In order to try to factor out the effect of error *type* on the kind of verbalization (and therefore on the coding of noticing quality) that occurs, errors could be put into functional groups based on what a sufficient explanation in a think-aloud might require. A researcher could then analyze differences in noticing quality for an error *type* and look at how the differences in noticing quality for a particular kind of error were related to corrections in revisions.

An attempt was made in the first study to compare three specific kinds of errors in order to investigate whether some were easier to correct than others and whether one condition facilitated correction more than the others. This did not involve comparing different qualities of noticing and their relation to correction *within* an error type as discussed above; rather, it involved comparing lexical, article, and preposition errors to each other with respect to overall percentages of changes in accuracy. These three kinds of errors were chosen as possibly representative of linguistic items that could more or less simply be learned (lexical items), items for which a system must be understood (articles), and items that might be in between those two extremes (prepositions).

Unfortunately, measuring changes in linguistic accuracy from the standpoint of individual error types was problematic for several reasons and had to be abandoned. First of all, a student might introduce new, unrelated errors of a certain type and then repeat them throughout the revision, making a simple comparison of error quantity from essay to revision impractical and misleading. For instance, one participant introduced 400% more article errors just by making one mistake (unrelated to any corrections that had been made) and then repeating it several times over the course of his revision. Equally, a student might be able to notice one overarching problem and then correct all the related errors at once. For example, a student might change several verbs from the past tense to the present after noticing one text cohesion problem. In the end, the fact that each participant had a different number of errors and a different distribution of error types made statistical analysis difficult. An initial (problematic) attempt to compare individual error types across condition can be seen in Table 15. If researchers do wish to analyze how differences in noticing quality for a particular kind of error are related to corrections in revisions, it will be necessary to keep these problems in mind and restrict the analysis to errors that are not repeated throughout a story.

TABLE 15
Percentages of correction for individual error types compared across condition
(in percentage form, problematic)

Error Type	Condition			Total
	EC	R	TA	
Prepositions (33)	57.58	40.99	28.03	45.84
Articles (35)	12.72	32.76	30.68	18.88
Lexical/phrase choice (25)	62.70	77.49	37.57	69.38

5.3 Further Research

Even without collecting any more data, additional analyses investigating a variety of questions could be performed. One thing to ask might be whether quality of noticing is related to level of L2 proficiency. Based on their exploratory study of two learners, Qi and Lapkin suggested that lower-proficiency L2 learners may not be able to notice the gap as well as higher-proficiency learners. With our data, we might be able to investigate this idea, comparing Level 300 IEP students with Level 093 EAP students. Collecting more data, we could also use technology to improve our power of observation. The possibilities of using videotapes and tracking eye movements provide many interesting avenues for future research. It might also be possible to manipulate time factors again and give all of the groups extra time to rehearse the language items they have detected.

In any case, in a future approach to the question of whether higher quality noticing leads to greater accuracy in revisions, it may be helpful to restrict the investigation to one condition (e.g., Think-Aloud) and even one kind of error. That way, it might be possible to focus on certain structures and design similar tasks or post-tests targeting the same linguistic forms. Post-tests could even be individualized as they were in Gass (1983) by keeping grammatical errors the same, but changing lexical items so that the participants would not recognize their writing. It will also be necessary to define “quality of noticing” better. The multiple distinctions used in this study (Tier II) were not very amenable to producing clear-cut quantitative results, and since we cannot assert that more elaborate verbalizations necessarily corresponded to deeper processing, the

substantive vs. perfunctory distinction used by Qi and Lapkin might not adequately take into account the gray area that exists between those extremes.

An important piece to the noticing puzzle might also come from focusing our attention on instances when participants have evidently noticed the gap. Qi and Lapkin found several exclamatory utterances in their think-alouds (e.g., “Oh! Yeah! Ha! I forgot this!”), and they took this to demonstrate that their participants were constantly engaging in comparisons of IL and TL. They also assumed that the participants’ original problems and experiences producing output influenced what they noticed while comparing their stories to reformulations. Controlling for error type, we could analyze our own data to find out if what we labelled as “noticing the gap” is correlated with corrections in revisions. If high quality noticing and noticing the gap are related to greater accuracy in revisions, teachers might be able to use this information to help their students process feedback.

5.4 Implications for Pedagogy

Even though it is not clear based on the results of this study whether the use of reformulations and think-alouds themselves can improve the quality of students’ noticing, it is possible to speculate that reformulation as a pedagogical technique might have some advantages over the way that explicit error correction is currently practiced that were not observable in this study. As has already been discussed, one of the main concerns with corrective feedback is that it often does not result in uptake. Practically speaking, this

makes sense. If teachers never set aside time for their students to sit down and evaluate their mistakes for the purposes of rewriting and incorporating suggestions, many students might simply glance at their grades, put their papers into their folders, and never look at them again. In that case, Zamel (1985) and Truscott (1996) have an even stronger position when they assert that error correction takes teachers' time away from other, more important aspects of students' writing.

What the current study suggests is that the simple act of having students make comparisons for 15 minutes might be somewhat effective in itself. It is possible that the *time* our participants devoted to comparison provided what was necessary to make error correction more useful. If a teacher's goals are to raise students' levels of awareness about common mistakes and to assist them in developing appropriate cognitive strategies, then perhaps error correction and reformulation can be utilized, not necessarily just as feedback on papers, but in an in-class activity designed to induce consciousness-raising and build more relationships between explicit and implicit knowledge. If students use noticing as a conscious cognitive process to focus attention on grammatical features that have given them trouble in output, their attempts at understanding them in input might be facilitative for acquisition.

As students improve in their abilities to make comparisons and notice differences, it might also be advantageous to elaborate on the process and exploit more of its potential benefits. One possibility would be to give learners reformulations or error corrections for a short period of time and then have them explain to each other in pairs the differences they have noticed. After that, a teacher might give them time to rewrite the essays in class and compare them again to see which changes they have been able to incorporate

and which ones they may have missed. This might serve to consolidate their knowledge and improve their accuracy in using the forms. It makes sense, as Fotos (1993) has pointed out, that in order to make the effects of consciousness-raising more durable, it is helpful to expose learners to the differences they have noticed more than once. In addition to this, the findings of this research seem to indicate that teachers should make changes as salient as possible (as was the case in the Error Correction condition) so that students find it easy to locate differences. Teachers should also make sure that the learners are given enough time to process and make use of the corrections (as was apparently not the case in the Reformulation and Think-Aloud conditions). Learners could also be encouraged to use teachers' corrections in order to perform error analyses and keep track of the errors they characteristically make.

Not only are the relationships between quality of noticing, feedback processing, interlanguage, and output theoretically interesting, but they are also very important for pedagogy. It would be extremely helpful to know what learners themselves are aware of as they compare two pieces of writing and then revise based on the insights they have gained. Research in the area of L2 learners' conscious cognitive processes and awareness will be especially fruitful if we can use it to help students develop effective strategies for noticing and maximize their ability to obtain intake from input.

APPENDICES

APPENDIX A

TABLE 16

Counterbalance Chart for Repeated Measures Study

Student	Time 1		Time 2		Time 3		Nationality
	cond.	pic.	cond.	pic.	cond.	pic.	
A	EC	A	R	B	T	C	Korean
B	R	C	T	A	EC	B	Japanese
C	T	A	EC	B	R	C	Korean
D	T	B	EC	C	R	A	Indonesian
E	R	B	T	C	EC	A	Korean
F	EC	A	R	B	T	C	Korean
G	R	C	T	A	EC	B	Japanese
H	T	A	EC	B	R	C	Korean
I	T	B	EC	C	R	A	Korean
J	T	C	EC	A	R	B	Korean
K	EC	C	R	A	T	B	Korean
L	EC	A	R	B	T	C	Japanese
M	EC	B	R	C	T	A	Korean
N	R	B	T	C	EC	A	Korean
O	R	A	T	B	EC	C	Korean

Condition:

EC = error correction

R = reformulation

T = think-aloud

Picture:

A = dinner party

B = jogging

C = bank robbers

APPENDIX B

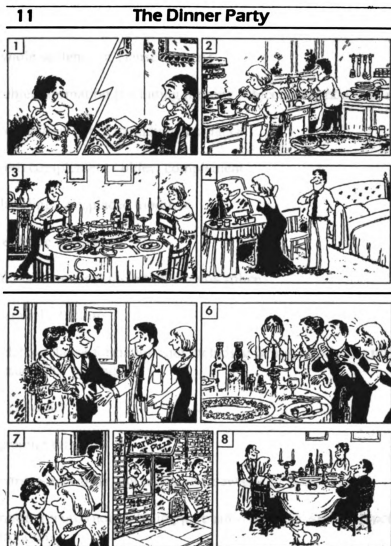


Figure 1. Writing Prompt A

All three picture sequences used in this study were adapted from the following source and used with the permission of the publisher.

Fuchs, M., Fletcher, M., Birt, D. (1986). *Around the World: Pictures for Practice, Book 2*. White Plains, NY: Longman, Inc.

(A) The Dinner Party, pp. 42-43 (B) Jogging, pp. 30-31 (C) Bank Robbers, pp. 14-15

APPENDIX C

Error Classification System

(adapted from Polio (1997), in turn adapted from Kroll (1990))

- 1 whole sentence or clause aberrant
- 2 subject formation (including missing subject/existential, but not wrong case)
- 3 verb missing (not including auxiliary)
- 4 verb complement / object complement
- 5 dangling / misplaced modifier
- 6 sentence fragment
- 7 run-on sentence (including comma splice)
- 8 parallel structure
- 9 relative clause formation (not including wrong or missing relative pronoun or resumptive pronoun)
- 10 word order
- 11 gapping error
- 12 extraneous words (not included elsewhere in descriptors)
- 13 missing word (not including preposition, article, verb, subject, relative pronoun)
- 14 wrong or extra modal
- 15 verb tense / aspect (incorrect tense, not incorrect formation)
- 16 voice (incorrect voice, not incorrect formation)
- 17 verb formation (including no auxiliary verb, lack of “to” with infinitive, participle misformation, gerund / infinitive problem)
- 18 subject-verb agreement

- 19 two-word verb (separation problem, incorrect particle)
- 20 noun-pronoun agreement (including wrong relative pronoun)
- 21 quantifier-noun agreement (much / many, this / these)
- 22 epenthetic pronoun (resumptive pronoun in relative clause, pronominal copy)
- 23 ambiguous or unlocatable reference; wrong pronoun
- 24 wrong case
- 25 lexical / phrase choice (including so / so that)
- 26 idiom
- 27 word form
- 28 wrong noun phrase morphology (but not word form)
- 29 wrong comparative formation
- 30 singular for plural
- 31 plural for singular
- 32 quantity words (few / a few, many kinds of, all / the whole)
- 33 preposition
- 34 genitive (missing / misused 's, N of N misuse)
- 35 article (missing, extra, incorrect)
- 36 deixis problem (this/that; the/this; it/that)
- 37 punctuation / mechanics (missing, extra, wrong; including restrictive / non-restrictive problem, capitalization, hyphens, indentation; *not* including commas after prepositional phrases)
- 38 negation (never/ever, any/some, either/neither, misplaced negator)
- 39 spelling (including not knowing the exact word, but attempting an approximation)
- 40 wrong or missing possessive

Notes:

- a.) If a sentence at the end of an essay is not finished, do not code it.
- b.) Code errors so that the sentence is changed minimally. If there are two possible errors requiring equal change, code the first error.
- c.) If tense is incorrect and misformed, count it as both 15 and 17. If there is a problem with both verb tense and subject-verb agreement, count it as both 15 and 18.
- d.) If an error can be classified as a relative clause error or a verb formation error, (ex: I know a man call John), count it only as verb formation.
- e.) Do not double-penalize for subject-verb agreement and a singular-plural problem (e.g., Visitor are pleased with the sight. (only a 30))
- f.) Count an error with quotation marks as only one error. Count a problem with a restrictive/non-restrictive relative clause as only one error.
- g.) Do not count the lack of a comma after an introductory prepositional phrase as an error.

APPENDIX D

In-Class Instructions

Today we are going to do a short pre-revising activity. I have typed copies of the stories that you wrote on Tuesday and made some changes to help you revise them. Each of you will receive a clean copy of your original story. Then, some of you will receive another copy of your story with writing *on* it, while others of you will receive a copy of your story that has been changed a little bit so that the writing sounds more native-like. If I don't give you any papers, you can read your novel for 15 minutes while the other students are working.

Please take 15 minutes right now to compare the two versions of your stories and try to find the differences. You can make marks on your papers if you like, but I will collect all of your papers at the end of class. Tomorrow, I will give you just a clean copy of your original story, and you will have 20 minutes to revise it.

APPENDIX E

An Example of Error Coding

(Student A, Error Correction Condition, Writing Prompt A)

One day, Mr. Smiths invites Mr. Kim at the dinner party at 8 P.M on Friday in his
house on the phone. On Friday, Mr. Smiths comes back home early for help his wife to
prepare the dinner. In the dinner party, the main food is going to be baked fish. His wife
cooks it and Mr. Smiths washes dishes. And they make the table together. After they
~~prepare~~ everything completely, they get dressed and wait for Mr. Kim and his wife. Mr.
Kim and his wife come to Mr. Smiths's house, and Mr. Smiths's couple receive them
friendly. As soon as all of them are in the dinning room, they notice that something
happend. The baked fish that was going to be main dish is disappeared. Mr. Smiths's
couple are embarrassed in that situation because they visited guests and prepared the
dinner party. Mr. Kim's couple also surprised that the main dish is gone. Everyone try to
pretend to be fine and nothing and Mr. Smiths who is host today run to Restaurant that
sells pizza and buy it. Finally they eat the pizza for the dinner party instead of the baked
fish. But nobody knows where the main dish is gone and who steals it. There is one who
knows the true. It is the Mr. Smiths's cat who is licking its paws.

An Example of Explicit Error Corrections

(Student A, Error Correction Condition, Writing Prompt A, Time 1)

One day ^{on the phone} Mr. Smiths invites Mr. Kim ^{to a} ~~at the~~ dinner party at 8 P.M. ^(at) on Friday in his house ~~on the phone~~. On Friday, Mr. Smiths comes back home early ^{to} ~~for~~ help his wife to prepare the dinner. ^{At} ~~In~~ the dinner party, the main food is going to be baked fish. His wife cooks it and Mr. Smiths washes ^(the) dishes. And they ^{set} ~~make~~ the table together. After they prepare everything completely, they get dressed and wait for Mr. Kim and his wife. Mr. Kim and his wife come to Mr. Smiths's house, and ^{the Smiths} ~~(Mr. Smiths's couple)~~ receive them cordially friendly. As soon as all of them are in the ^{dining} ~~dining~~ room, they notice that something has happened. The baked fish that was going to be ^{the} main dish ^{has} ~~is~~ disappeared. ^{the Smiths} ~~Mr. Smiths's~~ couple are embarrassed in that situation because they ^(this) ~~visited~~ ^{have invited} guests and prepared the dinner party. ^{The Kims are} ~~Mr. Kim's couple~~ also surprised that the main dish is gone. Everyone ^{tries} ~~try~~ to pretend to be fine ~~and nothing~~ and Mr. Smiths ^{the} who is ^{runs a restaurant} ~~host today runs to~~ Restaurant that sells pizza and ^{buys} ~~buys~~ it. Finally they eat the pizza for the dinner party instead of the baked fish. But nobody knows where the main dish ^{has} ~~is~~ gone and who ^{has stolen} ~~steals~~ it. There is one who knows the ^{truth} ~~true~~. It is ~~the~~ Mr. Smiths's cat who is licking its paws.

An Example of a Story and its Reformulation

(Student I, Reformulation Condition, Writing Prompt A, Time 3)

Story:

Smith who wear white sweater called his friend, Tom who wear black suit case for inviting dinner. Tom was glad to be invited by Smith. Tom memoed the appointmend, "8 p.m Friday Dinner with Smiths". Jane who is Smith's wife and Smith prepared Dinner for Tom and his wife. Smith and Jane prepared big fish for special menu. Smith and Jane almost done to set table, at that time Tom and his wife also almost done to wear good dress. Tom and his wife came to the Smith's house. They greeted gladly. They went to the table, and they see the food. However special menu which made by fish was gone. Smith and Jane were so embarrassed. Smith went to buy Pizza instead of Special menu. Other people waited for Smith sitting on the chairs. Where had the special menu gone? Smith and Jane's cat had eaten the special menu.

Reformulation:

Smith, who was wearing a white sweater, called his friend Tom, who was wearing a black suit, to invite him over for dinner. Tom was glad to have been invited by Smith. Tom wrote himself a memo about the appointment: "8 p.m. Friday, Dinner with Smiths." Smith and Jane, who is Smith's wife, prepared dinner for Tom and his wife. Smith and Jane prepared a big fish as a special menu.

Smith and Jane were almost done setting the table. At the same time, Tom and his wife were also almost done putting on good clothes. Tom and his wife arrived at the Smiths' house. They greeted each other gladly. Then they went to the table, and they saw the food. However, the special dish which was made with fish was gone. Smith and Jane were so embarrassed. Smith went to buy pizza to replace the special dish. The other people sat in chairs as they waited for Smith. Where had the special dish gone? Smith and Jane's cat had eaten the special dish.

TABLE 17

An Example of an Error Tally Sheet (Student A)

Error	Error Correction		Reformulation Think-Aloud					
	Time 1		Time 2		Time 3		Totals	
	Story	Revision	Story	Revision	Story	Revision	St.	Rev.
1					3	2	3	2
2			0	1	1	0	1	1
3			0	1	1	0	1	1
4								
5								
6								
7			1	1			1	1
8								
9								
10	1	0	4	2			5	2
11								
12	1	4	1	1			2	5
13					2	3	2	3
14								
15	3	0	8	8	4	1	15	9
16	1	1	1	1	2	2	4	4
17	2	2	4	1			6	3
18	3	0			1	1	4	1
19								
20								
21								
22								
23								
24								
25	6	1	12	3	1	3	19	7
26								
27	1	0	1	0	2	0	4	0
28								
29			1	0			1	0
30	0	3			10	2	10	5
31			2	2			2	2
32								
33	3	1	5	0	4	2	12	3
34								
35	6	4	5	5	8	9	19	18
36	1	0	1	0	1	1	3	1
37	5	4	4	0	4	2	13	6
38								
39	2	4	9	5	6	2	17	11
40								
totals	35	24	59	31	50	30	144	85

APPENDIX F

Think-Aloud Instructions

In order to help you revise your story tomorrow, I have typed 2 copies of it. This copy is the original version that you wrote [SHOW]. The other one [SHOW], I changed a little bit to make the writing sound more native-like. Soon, I will give you 15 minutes to compare the two copies and try to find the differences. You can make marks on the paper if you want to, but please also try to *talk out loud* as you compare the two versions. While you are doing this, I will use a tape recorder to record what is happening. It's not a test, so don't worry about being correct; pretend I'm not here and just say everything you're thinking as you compare the two essays – even if you're not sure. You don't have to talk in complete sentences, and don't worry about your grammar while you're speaking. Just talk about the differences you see. I won't talk at all or answer any questions. I'll just sit over here. Tomorrow, I will give you a clean copy of your original story, and you will have 20 minutes to revise it.

First, we will practice without the tape recorder so that you feel comfortable. Do you have any questions? (....)

Here is a story that another student wrote [SHOW] about this picture [SHOW], and here is a native speaker version [SHOW]. Please take about 5 minutes right now to compare them and talk about the differences that you see. (....)

(When finished practicing) OK, just remember to keep talking the whole time and say everything that goes through your head while you look at the two stories. Are you ready to start? I'll turn on the tape recorder now. Please start whenever you are ready.

APPENDIX G

An Example of Columns Format (Study 2, Student 13, Think-Aloud Condition)

Story	Reformulation
One day, he noticed that his tammy is kind of terrible by looking at the mirror.	One day, while looking in the mirror, a man noticed that his tummy looked pretty terrible.
Near by the mirror, there was a book titled "Get in Shape".	Nearby the mirror, there was a book entitled "Get in Shape."
He decided to start jogging.	He decided to start jogging.
He looked he was filled with a bunch of enagy.	He looked as though he was filled with a bunch of energy.
He read the book to know "how to jog".	He read the book in order to find out how to jog.
As he was jogging, his tammy was shaken,	As he was jogging, his tummy was shaking,
and his way of jogging was kind of strange.	and his jogging style was kind of strange.
everyone was pointing out him and laughing.	Everyone was pointing at him and laughing.
He was embarrassed.	He was embarrassed.
As soon as he turened the corner, he found two wealty females.	As soon as he turned the corner, he found two wealthy females.
One of them had a dog by holding a rope.	One of them was holding a dog on a leash.

Think-Aloud

OK, um, uh, I wrote, first of all, I wrote, he noticed that his tummy is kind of terrible, but native speaker's one is first of all, while looking in the mirror, a man noticed that his tummy looked pretty terrible, terrible. Mmm... I don't know why. I think... first of all, when I wrote this, I thought this is, I tried to write sentence... correctly, so I don't know why this is, why they, there is difference.

Hm. By looking at the mirror, and while looking at the mirror, while looking. Ah, and I also didn't know that when I used, when somebody uses the word while, I thought a person has to put sub- subject and verb and... but this time, she doesn't use any subject between while and looking. So... that's my, that's what I notice.

Mmmm... I wrote there was a book titled "Get in Shape," but another one's... there was a book entitled "Get in Shape." Hm. Maybe I should have wrote, written, entitled.

He looked he was filled with a bunch of energy. He looked he was filled... He was, he looked he was filled with a bunch of energy. That's what I wrote, and he looked as though he was filled with a bunch of energy... mmm. I didn't write "as though." Hm. Maybe if I wrote "as though" it's more, much more very, very more clear.

He read the book to know how to jog. He read the book in order to find out how to jog. He read the book to know how to jog. Hm. In order to find, find out. It's... makes more sense.

As he was jogging, his tummy was shaken, shaking. Hm.

Revision

One day, while he was watching the mirror, he noticed that his tammy was kind of terrible.

Near by the mirror, there was a book antitled "Get in Shape."

He decided to start jogging.

*He looked he was filled with a bunch of enagy.

He read the book in order to know "how to jog."

As he was jogging, his tammy was shaking,

APPENDIX H

Guidelines for Division into T-units (adapted from Polio, Fleck, & Leder, 1998)

- a.) A T-unit is defined as an independent clause and all its dependent clauses.
- b.) Count run-on sentences and comma splices as two T-units with an error in the first T-unit.

35 *awk* 7 17 25
Ex: The blood came out from his knee, the dog who got mad bited his wrenkle.
(*T-unit with 2 errors, 1 awk*) / (*T-unit with 2 errors*)

- c.) For sentence fragments, if the verb or copula is missing, count the sentence as 1 T-unit with an error. If an NP is standing alone, attach it to the preceding or following T-unit as appropriate and count it as an error. If a subordinate clause is standing alone, attach it to the preceding or following sentence and count it as an error.
- d.) When there is a grammatical subject deletion in a coordinate clause, count the entire sentence as 1 T-unit.
- e.) Count both “so” and “but” as coordinating conjunctions. Count “so that” as a subordinating conjunction unless “so” is obviously meant.
- f.) Do not count tag questions as separate T-units.
- g.) Count S-nodes with a deleted complementizer as a subordinate clause, as in: I believe that A and (that) B = 1 T-unit.
- h.) However, direct quotes should be counted as: John said, “A and B.”

1 T-unit / 1 T-unit

- i.) Assess the following types of structures on a case-by-case basis: If A, then B and C. As a result, A or B.
- j.) Count T-units in parentheses as individual T-units.

APPENDIX I

Coding System for Changes in Accuracy

Notes:

- a.) If 2 sentences are given as examples below, the first is the student's original version, and the second is the revised version. If 3 sentences are presented, the first is the original, the second is the reformulation, and the third is the revision.
- b.) An expression marked as awkward in the first sentence is considered an error. A new awkward expression in the revised sentence is not considered an error.

Original system:

1 error-free to error-free

They cook food, wash dishes, and clean the house.

They cook food, wash dishes, and clean the house.

2 error-free to error(s)

At 8:30, Mr. Crowley and his wife arrive at Smiths's house.

At 8:30, Crowleys' arrive at Smith's house.

3 error(s) to error-free

He could know many people laugh at him because of his looks.

He knew many people were laughing at him because of his looks.

4 error(s) to partial correction (but still not error-free)

He was compretly wety and bloody.

He was completely wet and bloody.

- 5 *error(s) to additional error(s) that are brand new and unrelated to those that were targeted*

It was a hard day to him.

It was a hard day for him.

It was hard day to him.

- 6 *error(s) to the same error(s) (no change)*

and they left too many evidence everywhere inside the bank.

and they left too much evidence everywhere inside the bank.

and they left too many evidence everywhere inside the bank.

- 7 *error(s) to different error(s) (attempted change of what was targeted, but no improvement)*

So, he made up his mind to make nice shape with jogging.

So, he made up his mind to improve his physique by jogging.

So, he made up his mind to make slim body with jogging.

- 3+5 *error(s) to error-free, except for a new, unrelated error*

During his wife talks with Mr. Crowley and his wife, Smiths goes out to
buy Pizza.

While his wife talks with Mr. Crowley and his wife, Smiths goes out to
buy pizza.

While his wife talks with Crowleys', Smith goes out to buy pizza.

3+7 *error(s) to error-free, except for an attempted change with no improvement*

One day, he get to know that he gains weight, watching himself on mirror.

One day, looking at himself in a mirror, he realizes that he has gained weight.

One day, looking at himself through the mirror, he realizes that he has gained weight.

4+5 *error(s) to partial correction of what was targeted, but also a new, unrelated error*

When they come to the dinner room. They really surprise because the fish has gone.

When they go to the dining room, they are really surprised because the fish is gone.

When they go to dinner room, they are really surprised because the fish is gone.

If you give me Marias Pizza, I can return your fish!!

If you give me Maria's Pizza, I can return your fish!

If you give me the Maria's Pizza, I can return the your fish!!

4+7 *error(s) to partial correction, plus an attempted change with no improvement*

He made his mind to do jogging every moring around his villiage.

He made up his mind to go jogging every morning around his village.

He made his mind to go jogging every morning around his villige.

6+7 *no changes except for an attempted change with no improvement*

However, when he really start “jogging”, he is very embarrassed.
However, when he really starts jogging, he is very embarrassed.
However, when he really start “jogging”, he is very embarrassed.

Revised System:

- + Evidence of noticing: At least one error in the T-unit was changed in the direction of the reformulation or correction.
(includes previous categories 4, 7, 3+7, 4+5, 4+7, 6+7)

- o Evidence of noticing: All the errors that existed in the original T-unit were completely corrected in the revised T-unit.
(includes previous categories 3, 3+5)

- No evidence of noticing: Nothing was changed.
(includes previous categories 5, 6)

- n/a Not applicable: The original T-unit did not contain an error, or the T-unit was added or deleted.
(includes previous categories 1, 2)

APPENDIX J

3-Tiered Coding System for the Quality of Noticing Related to Each Error, Based on Think-Aloud Data

TIER I

Whether or not each error was noticed, and whether or not it was changed (6 possibilities)

+ Noticing, + Correction	+ N	+ C
- Noticing, + Correction	- N	+ C
+ Noticing, - Correction	+ N	- C
- Noticing, - Correction	- N	- C
+ Noticing, + Change	+ N	+ H
- Noticing, + Change	- N	+ H

Interrater reliability:

99.44% for Noticing coding

98.05% for Correction/Change coding

TIER II

Quality of Noticing (Subcategory of + Noticing)

Interrater Reliability: 85.24%

M Mentioned only or read again with special emphasis

Student P: Oh, looked *at*, I missed 'at.'

SP Misspelling

Student Q: And they threat, threatened, I know this is wrong spell, so, yeah, change it. Mmm... threatened people in the bank with their guns.

ML Use of metalanguage without an explanatory reason

Student P: The women were upset, upset, ah, *with* him. Upset with him. I also confused, um, what kind of preposition I have to choose.

SM Stupid mistake

Student P: The women were upset with him because they were... right! They were worried! (laughs) Why I put 'worry' – yeah, right! Worried. They were worried about hurting her, her dog.

RE Reason

Student P: Oh, right, and. Yeah, I had to put 'and' because I want, I want to connect two sentences, so I have to... use a connecting word.

LN New lexical item

Student E: Oh, I learned a new vocabulary: 'make out.' Make out, make out means about maybe, mm, determine?

LO Old lexical item

Student E: Um, sometimes in my, in my worksheet, uh, I wrote down 'delightfully,' but the, the closer meaning is 'cheerfully,' so I... I change, I have to change 'delightfully' to 'cheerfully.'

NR Lack of reason

Student P: Unfortunately... unfortunately, it started to rain. Here I don't know why put the comma. (laughs) Actually, I, yeah, I don't know where I have to put comma or semi-colon. Actually, I'm, I'm every day confused.

RJ Rejection of change

(No examples available, but this would have been something like, "No, that's not what I meant to say.")

WR Wrong reason

Student E: I think the verb 'let' and verb 'make' is, uh, similar, so I... I wrote the 'let.' 'Let' and 'make' is, uh, si- same meanings sometimes, has a same meanings, but... uh, this situation, maybe 'make' is, uh, acceptable.

RD Reading the correction aloud

TIER III

Noticing the Gap

(Evidence of knowing that there was a problem in the first draft)

Student H: Oh, actually, I didn't know about the past verb of 'smell,' so I just write, wrote down the present verb, so I miss, missed.

Student I: And one of them was holding, holding a leash attached to a dog's neck. Actually, uh, it is hard to describe, describe the picture. Actually, I, I, I can't, I can't describe the picture, so I just put, put the words... Yeah, right. Holding a leash attached to a dog's, dog's neck. Actually, I don't know the word 'leash.' I don't know the word.

APPENDIX K

Selected Quotations from the Post-Study Debriefings

Student I "If I say something, my thinking is very fast because first I think and then I speak. But reading is, uh, I, I can think, think enough time. Yeah, and... um, and, yeah, I can think enough time, and I can think deeply. And sometimes I can memorize my mistakes, so last time I, I rewrite correct. But the think-aloud is, uh, actually, my problem is I, I forget everything easily. That is my problem. So I think this class activity is, uh, easy to memorize."

"When I speak English, I am very worry about, worry about grammar, so even though I find out my mistake, I, I have to, actually, it is, my mistake is not important in think-aloud because I concen- How can I explain to you? So... uh, even though I found a, my mistake, I... yeah, it is hard to memorize my mistake.

Experimenter: "Even though it wasn't like a conversation, even though it was just you talking?"

"Yeah, because my head is not good. My memorize is very bad, so... um, actually, this, this correction or comparing activity, I can memorize, I don't need to speak something, so I can memorize easily, but think-aloud is, uh, I, uh, notice my mistake, and then I have to do, tell you, and then I forgot."

"If I, this is maybe, this is language is Korean, maybe same. I also like to speak, uh, I mean, this is maybe Korean, Korean words, and this is Korean kind of Korean grammar class, and same situation, maybe I, um, I feel

more comfortable think aloud because, because, uh, actually, uh, in Korea, I study, study exam, the book is written by Korean, yeah, I like to go library, but sometimes I, I, I speak, I, yeah, I mean, this, this is Korean word, I read, read, and then I memorize, then I speak.”

Student J “I think correction is easiest, very, more easy than others because, in speech, when you speak for the first time, try to, it is very difficult to me because I have to speak why I, my sentence is wrong, so... I, I, I can understand why I... sentence is wrong, but I can’t speak well, yeah, so it is very difficult, and this also, compare, comparison activity also, I have to do searching why, what I... yeah, search, and so some, if I missed some words, I can, if I wrote some, missed some word, I, uh, just keep... uh, and correction activity, you write down, so, oh!, so I can search more easier, so I can’t... at, when I wrong word, if I, when I write down wrong word, oh! I can find it, so... yes.”

“I think first strategy I can’t count in, because I speak, I have to speak, so I think first is very difficult. I think I, it is important to me, I, when I, when I read that page, if I read, I can remember very well, but if I speech, I, I think it is more difficult to remember.”

Student F “I think correction activity is more understand. Easier, more easier. Because more familiar, I think, more familiar, when I watched this paper, I feel it’s more familiar. And... when I watched this paper, I felt, I recognized, this is wrong and this is right. I felt like that, so I think correction activity is more easy.”

“The most difficult thing is comparison because I, actually, I, I, um, that is very difficult to me, the compare, compare about, um, my case and the reviser case. It is difficult to distinguish. It is so difficult to distinguish.”

Student G “The correction is not useful for me because you already corrected... you used a different color pen or something, so I can find very easy to other mistake, ah, no, the difference. So... I just look faster to find color, Oh! I found it! Because it’s very easy to see a difference between my paper and (the one) you gave me. The paper makes it very easy to find the difference, so I don’t have to concentrate the paper.”

Student M “Talking is very hard, and... I don’t know grammar names. For example, I can speak a relative clause, but I don’t know a lot of grammar mistakes. So I can’t, I can’t explain my mistake.”

REFERENCES

REFERENCES

- Allwright, R.L., Woodley, M.P., & Allwright, J.M. (1988). Investigating reformulation as a practical strategy for the teaching of academic writing. *Applied Linguistics*, 9, 236-256.
- Baars, B. (1988). *A cognitive theory of consciousness*. New York: Cambridge University Press.
- Bialystok, E. (1994). Analysis and control in the development of second language proficiency. *SSLA*, 16, 157-168.
- Brinkman, J.A. (1993). Verbal protocol accuracy in fault diagnosis. *Ergonomics*, 36(11), 1381-1397.
- Cohen, A.D. (1987). Using verbal reports in research on language learning. In C. Faerch & G. Kasper (Eds.), *Introspection in second language research* (pp. 82-95). Philadelphia: Multilingual Matters, Ltd.
- Cohen, A. D., & Olshtain, E. (1993). The production of speech acts by EFL learners. *TESOL Quarterly*, 27(1), 33-56.
- Cumming, A. (1990). Metalinguistic and ideational thinking in second language composing. *Written Communication*, 7(4), 482-511.
- Dechert, H.W. (1987). Analysing language processing through verbal protocols. In C. Faerch & G. Kasper (Eds.), *Introspection in second language research* (pp. 96-112). Philadelphia: Multilingual Matters, Ltd.
- Doughty, C., & Williams, J. (Eds.) (1998). *Focus on form in classroom second language acquisition*. Cambridge: Cambridge University Press.
- Ellis, R. (1995). Interpretation tasks for grammar teaching. *TESOL Quarterly*, 29, 87-105.
- Ellis, R. (2001). Introduction: Investigating form-focused instruction. *Language Learning*, 51, SUPP/1, 1-46.
- Ericsson, K.A., & Simon, H.A. (1993). *Protocol analysis: Verbal reports as data (revised edition)*. Cambridge, MA: The MIT Press.
- Ericsson, K.A., & Simon, H.A. (1987). Verbal reports on thinking. In C. Faerch & G. Kasper (Eds.), *Introspection in second language research* (pp. 24-53). Philadelphia: Multilingual Matters, Ltd.

- Faerch, C., & Kasper, G. (1987). From product to process – Introspective methods in second language research. In C. Faerch & G. Kasper (Eds.), *Introspection in second language research* (pp. 5-23). Philadelphia: Multilingual Matters, Ltd.
- Ferris, D. (1995). Teaching students to self-edit. *TESOL Journal*, (summer), 18-22.
- Ferris, D. (1999). The case for grammar correction in L2 writing classes: A Response to Truscott (1996). *Journal of Second Language Writing*, 8(1), 1-11.
- Fotos, S.S. (1993). Consciousness raising and noticing through focus on form: Grammar task performance versus formal instruction. *Applied Linguistics*, 14(4), 383-407.
- Frodesen, J. (2001). Grammar in writing. In M. Celce-Murcia (Ed.), *Teaching English as a second or foreign language* (pp. 233-248). New York: Newbury House.
- Fuchs, M., Fletcher, M., & Birt, D. (1986). *Around the world: Pictures for practice, book 2*. White Plains, NY: Longman. pp. 30-31.
- Gass, S. (1983). The development of L2 intuitions. *TESOL Quarterly*, 17 (2), 273-291.
- Gass, S., & Mackey, A. (2000). *Stimulated recall methodology in second language research*. Mahwah, New Jersey: Lawrence Erlbaum Associates, Publishers.
- Grotjahn, R. (1987). On the methodological basis of introspective methods. In C. Faerch & G. Kasper (Eds.), *Introspection in second language research* (pp. 54-81). Philadelphia: Multilingual Matters, Ltd.
- Hacker, D.J., Plumb, C., Butterfield, E.C., Quathamer, D, & Heineken, E. (1994). Text revision: Detection and correction of errors. *Journal of Educational Psychology*, 86(1), 65-78.
- Hatch, E., & Lazaraton, A. (1991). *The research manual: Design and statistics for applied linguistics*. New York: Newbury House Publishers.
- Hauser, E. (2002). Incomplete verbalization in concurrent think-aloud protocols. Paper presented at the *Second Language Research Forum*, Toronto, Canada.
- Hayes, J.R., & Flower, L.S.J. (1983). Uncovering cognitive processes in writing: An introduction to protocol analysis. In P. Mosenthal, L. Tamor, and S.A. Walmsley (Eds.) *Research on writing: Principles and methods* (pp. 206-219). New York: Longman.
- Johnson, K. (1988). Mistake correction. *ELT Journal*, 42(2), 89-96.

- Jourdenais, R. (2001). Cognition, instruction and protocol analysis. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 354-375). Cambridge, UK: Cambridge University Press.
- Klein, W. (1986). *Second language acquisition*. Cambridge: Cambridge University Press.
- Laufer, B., & Hulstijn, J. (2001). Incidental vocabulary acquisition in a second language: The construct of task-induced involvement. *Applied Linguistics*, 22, 1 - 26.
- Leow, R.P. (2003). Awareness, different learning conditions, and L2 development. Paper presented at the *AAAL Annual Conference*, Arlington, Virginia.
- Leow, R.P. (1997). Attention, awareness, and foreign language behavior. *Language Learning*, 47, 467-506.
- Long, M. (1998). Focus on form in task-based language teaching. *University of Hawai'i Working Papers in ESL*, 16, 35-49.
- Mackey, A., Gass, S., & McDonough, K. (2000). How do learners perceive interactional feedback? *Studies in Second Language Acquisition*, 22, 471-497.
- Makino, T. (1993). Learner self-correction in EFL written compositions. *ELT Journal*, 47(4), 337-341.
- Nisbett, R.E., & Wilson, T.D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, 84 (3), 231-255.
- Norris, J. M., & Ortega, L. (2000). Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language Learning*, 50, 417-528.
- O'Malley, J., & Chamot, A. (1990). *Learning strategies in second language acquisition*. Cambridge: Cambridge University Press.
- Polio, C. (1997). Measures of linguistic accuracy in second language writing research. *Language Learning*, 47(1), 101-143.
- Polio, C., Fleck, C., & Leder, N. (1998). "If only I had more time:" ESL learners' changes in linguistic accuracy on essay revisions. *Journal of Second Language Writing*, 7(1), 43-68.
- Qi, D.S., & Lapkin, S. (2001). Exploring the role of noticing in a three-stage second language writing task. *Journal of Second Language Writing*, 10 (2001), 277-303.
- Robb, T., Ross, S., & Shortreed, I. (1986). Salience of feedback on error and its effect on EFL writing quality. *TESOL Quarterly*, 20(1), 83-95.

- Robinson, P. (1995). Attention, memory, and the "Noticing" Hypothesis. *Language Learning*, 45, 283-331.
- Russo, J.E., Johnson, E.J., & Stephens, D.L. (1989). The validity of verbal protocols. *Memory and Cognition*, 17 (6), 759-769.
- Schmidt, R.W. (1990). The role of consciousness in second language learning. *Applied Linguistics*, 11(2), 129-158.
- Schmidt, R., & Frota, S. (1986). Developing basic conversational ability in a second language: A case study of an adult learner of Portuguese. In R. Day (Ed.), *Talking to learn*, 237-326.
- Smagorinsky, P. (1989). The reliability and validity of protocol analysis. *Written Communication*, 6 (4), 463-479.
- Smagorinsky, P. (1994). Think-aloud protocol analysis: Beyond the black box. In P. Smagorinsky (Ed.), *Speaking about writing: Reflections on research methodology* (pp. 3-19). Thousand Oaks, CA: Sage.
- Steinberg, E.R. (1986). Protocols, retrospective reports, and the stream of consciousness. *College English*, 48 (7), 697-712.
- Stratman, J.F., & Hamp-Lyons, L. (1994). Reactivity in concurrent think-aloud protocols: Issues for research. In P. Smagorinsky (Ed.), *Speaking about writing: Reflections on research methodology* (pp. 89-112). Thousand Oaks, CA: Sage.
- Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. Gass & C. Madden (Eds.), *Input and second language acquisition*. (pp. 235-256). Rowley, MA: Newbury House.
- Swain, M. (1995). Three functions of output in second language learning. In G. Cook & B. Seidhofer (Eds.), *Principles and practice in applied linguistics* (pp. 125-144). Oxford: Oxford University Press.
- Swain, M., & Lapkin, S. (1995). Problems in output and the cognitive processes they generate: A step towards second language learning. *Applied Linguistics*, 16 (3), 371-391.
- Thornbury, S. (1997). Reformulation and reconstruction: Tasks that promote "noticing". *ELT Journal*, 51, 326-335.
- Toms, M. (1992). Verbal protocols: How useful are they to cognitive ergonomists? In E.J. Lovesey (Ed.), *Contemporary ergonomics*. Proceedings of the Ergonomics Society's 1992 Annual Conference (Taylor & Francis), 316-321.

- Truscott, J. (1996). The case against grammar correction in L2 writing classes. *Language Learning*, 46 (2), 327-369.
- Truscott, J. (1998). Noticing in second language acquisition: A critical review. *Second Language Research*, 14, 103-135.
- Truscott, J. (1999). The case for "The case against grammar correction in L2 writing classes": A response to Ferris. *Journal of Second Language Writing*, 8(2), 111-122.
- Williams, A.M., & Davids, K. (1997). Assessing cue usage in performance contexts: A comparison between eye-movement and concurrent verbal report methods. *Behavior Research Methods, Instruments, & Computers*, 29(3), 364-375.
- Zamel, V. (1985). Responding to student writing. *TESOL Quarterly*, 19(1), 79-101.

MICHIGAN STATE UNIVERSITY LIBRARIES



3 1293 02461 4525