WRITTEN FEEDBACK IN SECOND LANGUAGE ACQUISITION:
EXPLORING THE ROLES OF TYPE OF FEEDBACK, LINGUISTIC TARGETS, AWARENESS,
AND CONCURRENT VERBALIZATION

A Dissertation
submitted to the Faculty of the
Graduate School of Arts and Sciences
of Georgetown University
in partial fulfillment of the requirements for the
degree of
Doctor of Philosophy
in Linguistics

By

Bo Ram Suh, M.A.

Washington, DC
February 25, 2010
Copyright 2010 by Bo Ram Suh
All Rights Reserved
WRITTEN FEEDBACK IN SECOND LANGUAGE ACQUISITION:
EXPLORING THE ROLES OF TYPE OF FEEDBACK, LINGUISTIC TARGETS,
AWARENESS, AND CONCURRENT VERBALIZATION

Bo Ram Suh, M.A.

Thesis Advisor: Alison Mackey, Ph.D.

ABSTRACT

Very few studies have shown that written feedback facilitates L2 learning (although see Bitchener, 2008, Bitchener & Knoch, 2008, and Sheen, 2007), and studies exploring the relative effectiveness of different types of written feedback, in particular direct and indirect feedback, have yielded mixed results. Although the existing studies were premised on attention and awareness, learners’ attential processes have rarely been measured directly. This dissertation investigates the effects of type of written feedback on L2 development in relation to learners’ attential processes. Relationships between feedback type, linguistic structures, awareness, and L2 development are also considered, together with the issue of the potential reactivity of concurrent verbalization.

The study used a pretest-immediate posttest-delayed posttest design, with three treatment sessions. Eighty-one Korean university EFL learners, randomly assigned to one of five groups, completed three written story retellings guided by picture prompts. Participants assigned to the experimental groups differed as to: (1) what type of written feedback they received, and (2) whether they were asked to
think aloud during the feedback session. Participants in the control group did not receive any feedback. The targeted structures were the past counterfactual conditional and the objective-of-preposition type of relative clauses in English. In addition to tests of target structure recognition/interpretation, written story retelling tasks served as production tests to assess L2 development. Awareness was measured by both concurrent verbalization and a post-exposure questionnaire.

Results indicated that type of written feedback impacted participants’ development of one of the target structures, the past counterfactual conditional, with those participants who received direct feedback improving significantly immediately and one week after the treatments. Participants who received direct feedback also reported a higher level of awareness than those who received indirect feedback. These higher levels of awareness were associated with development, regardless of the type of linguistic target. Finally, concurrent verbalization had no effect on development in this investigation. These findings have theoretical, methodological, and pedagogical implications.
TABLE OF CONTENTS

LIST OF FIGURES ........................................................................................................ ix

LIST OF TABLES ......................................................................................................... xi

CHAPTER 1: INTRODUCTION ................................................................................ 1

CHAPTER 2: REVIEW OF RELATED LITERATURE .............................................. 9

2.1. Written Feedback and Instructed SLA .............................................................. 10
  2.1.1. Written Feedback and L2 Learning: Theoretical Background ............... 10
  2.1.2. Empirical Studies on the Effects of Written Feedback on L2 Learning . 12
    2.1.2.1. Empirical Findings on Effectiveness of Written Feedback .......... 12
    2.1.2.2. Empirical Findings on Relative Effectiveness of Different Types of
             Written Feedback .............................................................................. 22
      2.1.2.2.1. Empirical findings on relative effectiveness of direct and
                   indirect feedback ........................................................................ 23
      2.1.2.2.2. Empirical findings on relative effectiveness of different types of
                   indirect feedback ........................................................................ 27
      2.1.2.2.3. Empirical findings on relative effectiveness of different types of
                   direct feedback .............................................................................. 28
  2.1.3. Remaining Issues .................................................................................... 33
    2.1.3.1. Research Design and Methodological Issues ................................ 33
      2.1.3.1.1. Measurement of learning. ......................................................... 33
      2.1.3.1.2. Control group ........................................................................ 34
      2.1.3.1.3. Confounding variables .............................................................. 35
    2.1.3.2. Possible Role of Type of Linguistic Target in Mediating Written
             Feedback and L2 Development ............................................................. 36
2.1.3.3. Learner-internal Factor Linking Written Feedback and L2 Learning

.................................................................................................................................................. 38

2.2. Attention, Awareness, and SLA........................................................................................... 40
2.2.1. Theoretical Models of Attention and Awareness in SLA .............................................. 40
2.2.2. Empirical Studies on the Role of Awareness in SLA ..................................................... 44
2.2.3. Operationalization and Measures of Attention and Awareness in SLA ....................... 47
2.2.3.1. Concurrent Verbal Reports and Reactivity Issue ....................................................... 51
2.2.3.2. Empirical Studies on Reactivity in SLA ................................................................. 53

2.3. Rationale and Research Questions ................................................................................... 62

CHAPTER 3: RESEARCH DESIGN AND METHODOLOGY ....................................................... 65

3.1. Research Design.................................................................................................................... 65
3.2. Participants .......................................................................................................................... 66
3.3. The Targeted Linguistic Structures .................................................................................... 69
3.4. Operationalizations of Direct and Indirect Written Feedback .......................................... 73
3.5. Materials ............................................................................................................................... 74
3.5.1. Treatment Materials ........................................................................................................ 74
3.5.2. Assessment Materials....................................................................................................... 76
3.5.2.1. Recognition/Interpretation Test ................................................................................. 77
3.5.2.2. Production Test .......................................................................................................... 78
3.5.2.3. Measures of Awareness ............................................................................................ 78
3.5.2.3.1. Think-aloud protocols ......................................................................................... 79
3.5.2.3.2. Debriefing questionnaire ...................................................................................... 80
3.6. Procedure ............................................................................................................................. 80

3.7. Scoring Procedures ............................................................................................................ 85
3.7.1. Recognition/Interpretation Tests ............................................................. 85
3.7.2. Production Tests ...................................................................................... 86
3.7.3. Coding of Think-aloud Data ................................................................... 86
3.7.4. Complementary Use of Debriefing Questionnaire ................................. 92
3.8. Statistical Procedures ..................................................................................... 92

CHAPTER 4: RESULTS .......................................................................................... 95
4.1. Learning: Results for the Effects of Written Feedback Condition .......... 95
   4.1.1. Learning of Past Counterfactual Conditionals ........................................ 96
      4.1.1.1. Results on Recognition Task ........................................................... 96
      4.1.1.2. Results on Production Task ........................................................... 102
   4.1.2. Learning of OPREP Types of Relative Clauses ................................... 106
      4.1.2.1. Results on Recognition Task ......................................................... 106
      4.1.2.2. Results on Production Task ........................................................... 108
   4.1.3. Summary ............................................................................................... 111
4.2. Awareness: Results for the Effects of Type of Feedback on Reported Awareness ........................................................................................................... 112
4.3. Awareness and Learning: Results on the Effects of Reported Awareness .. 116
   4.3.1. Awareness and Learning: Results for the Past Counterfactual Conditionals .......................................................... 117
      4.3.1.1. Results on Production Task ........................................................... 117
   4.3.2. Awareness and Learning: Results for the OPREP Types of Relative Clauses ........................................................................................................... 120
      4.3.2.1. Results on Production Task ........................................................... 120
   4.3.3. Summary ............................................................................................... 123
4.4. Reactivity: Results for the Effects of Verbalization Condition ............... 123
CHAPTER 5: DISCUSSION AND CONCLUSION ............................................. 133

5.1. Discussion .................................................................................................... 133
  5.1.1. Effects of Type of Written Feedback on L2 Learning ......................... 133
  5.1.2. Effects of Type of Written Feedback on Reported Awareness .......... 139
  5.1.3. Link between Reported Awareness and L2 Learning ....................... 143
  5.1.4. Reactivity of Online Verbal Reports .................................................. 145

5.2. Conclusion ................................................................................................... 146
  5.2.1. Implications for Theory, Pedagogy, and Methodology ....................... 146
  5.2.2. Limitations and Suggestions for Future Research ............................... 149
  5.2.3. Conclusion .......................................................................................... 153

APENDIX A ........................................................................................................... 156

APENDIX B ........................................................................................................... 158

APPENDIX C ......................................................................................................... 161

APPENDIX D ......................................................................................................... 207

BIBLIOGRAPHY ................................................................................................... 213
LIST OF FIGURES

Figure 1. Procedure .................................................................................................. 83

Figure 2. Mean Scores on Recognition of Past Conditionals by Feedback Condition ........................................................................................................................... 98

Figure 3. Mean Scores on Production of Past Conditionals by Feedback Condition ........................................................................................................................................... 104

Figure 4. Mean Scores on Interpretation of Relative Clauses by Feedback Condition ......................................................................................................................................... 107

Figure 5. Mean Scores on Production of Relative Clauses by Feedback Condition ........................................................................................................................................... 110

Figure 6. Mean Scores on Production of Past Conditionals by Reported Awareness ........................................................................................................................................... 118

Figure 7. Mean Scores on Production of Relative Clauses by Reported Awareness ........................................................................................................................................... 121

Figure 8. Mean Scores on Recognition of Past Conditionals by Verbalization Condition ........................................................................................................................................... 129

Figure 9. Mean Scores on Interpretation of Relative Clauses by Verbalization Condition ........................................................................................................................................... 130

Figure 10. Mean Scores on Production of Past Conditionals by Verbalization Condition ........................................................................................................................................... 131
Figure 11. Mean Scores on Production of Relative Clauses by Verbalization Condition.......................................................... 132
LIST OF TABLES

Table 1 Number of Attempts at Target Sentences and Number of Feedback Received in Each Feedback Condition ................................................................. 85
Table 2 Examples of Think-aloud Data Coding Categories ............................................. 90
Table 3 Recognition of Past Conditionals: Mean Test Scores by Feedback Condition across Time ....................................................................................................... 96
Table 4 Recognition of Past Conditionals: 3 x 3 ANOVA for Feedback Condition by Time ........................................................................................................................................ 99
Table 5 Recognition of Past Conditionals: 2 x 3 ANOVA for Feedback Condition by Time ......................................................................................................................................... 101
Table 6 Production of Past Conditionals: Mean Test Scores by Feedback Condition across Time ........................................................................................................................................... 102
Table 7 Production of Past Conditionals: ANOVA for Feedback Condition by Time ................................................................................................................................................. 103
Table 8 Interpretation of Relative Clauses: Mean Test Scores by Feedback Condition across Time ........................................................................................................................................... 106
Table 9 Interpretation of Relative Clauses: ANOVA for Feedback Condition by Time ................................................................................................................................................. 108
Table 10 Production of Relative Clauses: Mean Test Scores by Feedback Condition across Time ........................................................................................................................................... 109
Table 11 Production of Relative Clauses: ANOVA for Feedback Condition by Time ................................................................................................................................................. 111
Table 12 Chi-square analysis for Type of Feedback by Level of Reported Awareness according to Targeted Linguistic Structure .............................................................. 114
Table 13 Production of Past Conditionals: Mean Test Scores by Reported Awareness across Time .............................................................................................................. 117
Table 14 Production of Past Conditionals: ANOVA for Reported Awareness by Time ................................................................................................................ 119
Table 15 Production of Relative Clauses: Mean Test Scores by Reported Awareness across Time ..................................................................................................... 120
Table 16 Production of Relative Clauses: ANOVA for Reported Awareness by Time ......................................................................................................................... 122
Table 17 Mean Test Scores by Verbalization Condition across Time .................... 125
Table 18 Reactivity of Concurrent Verbalization: ANOVA for Verbalization Condition by Time ....................................................................................................................... 127
CHAPTER 1: INTRODUCTION

A great deal of research in the field of second language acquisition (SLA) to date has investigated the role of feedback in L2 learners (e.g., Ashwell, 2000; Carroll & Swain, 1993; Carroll, Swain, & Roberge, 1992; Chandler, 2003; DeKeyser, 1993; Doughty & Varela, 1998; Ferris & Roberts, 2001; Fazio, 2001; Kepner, 1991; Iwashita, 2003; Lightbown & Spada, 1990; Long, Inagaki, & Ortega, 1998; Mackey, 1999; Mackey & Philp, 1998; Nagata, 1993, 1997; Nagata & Swisher, 1995; Polio, Fleck, & Leder, 1998; Robb, Ross, & Shortreed, 1986; Spada & Lightbown, 1993), and lends support for the overall facilitative effects of oral feedback on L2 development (e.g., Carroll & Swain, 1993; Iwashita, 2003; Leeman, 2003; Lightbown & Spada, 1990; Long, Inagaki, & Ortega, 1998; Mackey, 1999; Mackey & Philp, 1998; Spada & Lightbown, 1993). A positive association between oral feedback and L2 learning has also been supported by recent meta-analyses by Mackey and Goo (2007) and Russell and Spada (2006). However, the role of written feedback has been relatively under-investigated, and very few studies have shown that written feedback or error corrections on L2 learners’ written production, in fact, facilitate language learning (Bitchener, 2008; Bitchener & Knoch, 2008; Sheen, 2007).

While most written feedback research has compared different types of feedback on written errors in learners’ compositions in order to explore the optimal means of error correction, no one type of written feedback has been conclusively
shown to be more effective than another (e.g., Fazio, 2001; Lalande, 1982; Robb et al., 1986; Sheen, 2007). Written feedback has been characterized by different terminology and has been further defined through various distinctions, including direct or indirect, explicit, implicit, and coded or uncoded. One distinction that has been made in most of the written feedback research is between direct (i.e., errors are corrected) and indirect feedback (i.e., errors are not corrected, but are indicated through either underlining, circling, highlighting, or key table or checklist indicating error type) (e.g., Bates, Lane, & Lange, 1993; Ferris, 1995; Ferris & Hedgcock, 1998; Hendrickson, 1978, 1980; Lalande, 1982; Walz, 1982). However, as will become clear in the review of literature, regarding the relative efficacy of indirect and direct feedback, no definitive conclusions can be made to date, which begs further research.

The rationale of and directions for the present study are partly based on the research design and methodological limitations of prior empirical research on the developmental benefits of written feedback. Most noticeably, the focus of previous empirical research on written feedback has often been L2 learners’ performance on revision tasks of the same essay rather than learning outcomes. In other words, much of the research that investigated the effects of written feedback on the development of linguistic accuracy or grammatical form in learner production measured effectiveness in terms of improvement from the original to the revised productions. Performance research cannot provide a direct answer to whether written feedback
facilitates L2 learning. Furthermore, the frequent lack of a true control group makes it difficult to document the effectiveness of written feedback in terms of linguistic accuracy. There are also confounding variables that may be affecting previous studies’ results, which limits the accuracy of any observation of the isolated effects of written feedback. The potential confounds include (i) provision of supplemental grammar instruction, (ii) student-teacher conferences as part of the treatment, (iii) uncontrolled, possible aid of external resources (e.g., textbooks, dictionaries), and (iv) use of homework essay assignments as measurement of improvement. These issues are taken into account in designing and conducting the current investigation.

Furthermore, given the findings from studies in other strands of SLA research, one possible explanation for the mixed results of previous research may relate to the type of linguistic targets (e.g., Iwashita, 2003; Leow, Egi, Nuevo & Tsai, 2003; Long, Inagaki & Ortega, 1998; Mackey, Gass, & McDonough, 2000; Mackey & Philp, 1998; Nagata, 1993; Nagata & Swisher, 1995; Shook, 1994; VanPatten, 1990). That is, the effectiveness of written feedback may depend on the type of linguistic targets as has been shown to be the case in previous work on oral feedback, textual input enhancement, and computer-assisted language learning. The current dissertation research explores this possibility, which has scarcely been addressed in the research on written feedback.

In written feedback studies, it was assumed that written feedback provided on learners’ written compositions would draw learners’ attention to the features
marked/corrected, thereby promoting superior, subsequent awareness and further
cognitive processing of the attended forms, when compared to a no feedback
condition. However, many studies in this area only made an indirect inference
regarding the presence and the focus of attention and awareness as well as learners’
subsequent processing of L2 data by means of post-exposure tasks. It is of
theoretical and pedagogical interest to address whether empirical findings would
support or contradict the assumption that written feedback promotes subsequent L2
processing and learning in SLA. Although the operationalization of learners’
attentional processes is a thorny question, as acknowledged in the SLA literature
(e.g., it is crucial to the internal validity of research premised on the significant role
of attention in SLA to operationalize and measure them in a more direct way (Gass
& Mackey, 2000, 2007; Leow, 1999b; Mackey & Gass, 2005). Furthermore, recent
SLA studies have yielded positive evidence for the benefits of measuring learners’
attentional processes directly via think-aloud protocols (e.g., Leow, 1997, 1998a,
1998b, 2001a, 2001b; Leow et al., 2003; Rosa & Leow, 2004a, 2004b; Rosa &

In terms of issues of attention and awareness, while there is general
agreement that attention is a necessary condition for learning in both the fields of
cognitive psychology and SLA (e.g., Anderson, 1983; Dulany, Carlson & Dewey,
1984, 1985; Ellis, 1993; Ericsson & Simon, 1984; Long, 1991; Nissen & Bullemer,
1995, 2001; Sharwood Smith, 1981, 1986, 1993; Shiffrin & Schneider, 1977; Tomlin & Villa, 1994), the role of awareness in L2 development (i.e., whether attention necessarily includes awareness for subsequent processing of input to occur) has been an issue of controversy. Put simply, some SLA researchers have proposed that awareness (at the level of noticing) is a necessary condition for L2 learning to take place (e.g., Robinson, 1995, 1996, 2003; Schmidt, 1990, 1993, 1994a, 1994b, 1995), while others have posited that it is possible to learn without awareness, and it is detection that is necessary for subsequent learning (e.g., Tomlin & Villa, 1994).

Notwithstanding controversies in the literature over the necessity of awareness for L2 learning (Robinson 1995; Schmidt 1990; 1995; 2001; Tomlin and Villa 1994), empirical support supports a relationship between awareness and L2 learning, with higher levels of awareness more strongly associated with learning (e.g., Leow, 1997; Rosa & Leow, 2004; Rosa & O’Neill 1999). In this study, introspective methods of exploring awareness and attention may offer important insights on learners’ cognitive processes when learners are exposed to different types of written feedback on a variety of linguistic targets. By gathering more direct evidence on how those processes work and how they relate to L2 development, this study elucidates the relationship between factors such as type of written feedback, type of the targeted structures, learner awareness, and L2 learning.

At the same time, when using concurrent verbal reports as a research tool, the issue of reactivity needs to be considered in SLA. It has been cautioned, in both
L1 and L2 research, that concurrent verbalization might change learners’ cognitive processes and either positively or negatively affect their performance on language-related tasks (e.g., Jourdenais, 2001; Sachs & Polio, 2007; Stratmand & Hamp-Lyons, 1994). To date, six published SLA studies have empirically addressed this point. Five of them found no reactivity between silent and verbalization groups (Bowles, 2008; Bowles & Leow, 2005, Leow & Morgan-Short, 2004; Sachs & Suh, 2007; Sanz et al., 2008 (Exp.1)). A difference, however, was found between metalinguistic and non-metalinguistic verbalization groups (Bowles & Leow, 2005), and two found think-alouds to be reactive when compared to the silent group (Sachs & Polio, 2007; Sanz et al., 2008 (Exp.2)). A difference was also found between metalinguistic verbalization and silent groups in terms of item-learning (Bowles, 2008). The small, inconsistent body of research does not lend itself toward conclusive results due to the potential variables that may influence the reactivity, such as task type, type of measurement tests of L2 learning, and complexity or salience of targeted linguistic structure. Further research on this issue is thus clearly necessary for the field of SLA.

In summary, drawing upon the previous research in SLA on written feedback and attention and awareness, this dissertation research aims to investigate the developmental benefits of two types of written feedback in relation to learners’ attentional processes during the interaction with the feedback. An additional goal is to shed some light on the possible links between type of written feedback, learner
awareness, type of linguistic structure, and L2 learning. The study also addresses the potential reactive effects of concurrent verbal protocols on L2 development.

The organization of the dissertation is as follows: Chapter 2 includes a review of relevant literature. The first part of chapter 2 discusses the instructional benefits of written feedback on L2 learning. The chapter first summarizes previous research on the role of feedback in SLA, with a focus on the scant and inconsistent evidence for the role of written feedback. Chapter 2 reviews empirical studies that investigate the effects of written feedback on L2 learning and the relative effectiveness of different types of written feedback. The chapter then identifies theoretical and methodological issues that remain unsolved in the written feedback literature. The second part of chapter 2 discusses the role of attention and awareness in SLA as a learner-internal factor. This part of the chapter first introduces the theoretical constructs of attention and awareness and reviews empirical studies that have investigated the role of awareness in SLA. Then, the chapter addresses some methodological issues concerning the measurement of awareness. Chapter 3 presents specific research questions that guided the present empirical study, and introduce the research design together with a detailed description of the participants, materials, procedures, and statistical analyses used to investigate the research questions. Chapter 4 presents the results of the experiment by reporting both descriptive and inferential statistical analyses of data for each research question. Finally, chapter 5 includes a discussion of the results in light of both the research questions and the
empirical evidence available in the SLA field. The chapter concludes with a careful
discussion of the limitations of the current study and makes some preliminary
suggestions for further research.
CHAPTER 2: REVIEW OF RELATED LITERATURE

This chapter first provides a summary of previous research on feedback in SLA, highlighting a controversial issue on written feedback. Next, it reviews empirical investigations into the effect of written feedback on L2 learning as well as the relative effectiveness of different types of written feedback. Then, the research design and methodological issues that need to be considered for further research are discussed. It is followed by a discussion of two variables that may play a role in linking written feedback and L2 learning, but have not yet been addressed in previous research on written feedback: type of linguistic target, and attention and awareness as a learner-internal factor. In the second part of this chapter, the theoretical constructs of attention and awareness are first introduced with a review of the theoretical models and hypotheses that have been put forth in the field of SLA. Next, empirical studies on the role of awareness in L2 learning are reviewed. Then, measurements of the constructs employed in SLA research are reviewed including concurrent verbalizations and retrospective post-exposure questionnaires that pertain to the present study. As one of the validity issues of concurrent verbal reports, the potential reactivity is discussed and studies addressing the reactivity issue in SLA are reviewed.
2.1. Written Feedback and Instructed SLA

2.1.1. Written Feedback and L2 Learning: Theoretical Background

In the field of SLA, feedback, in general terms, refers to “information learners receive in response to their communicative efforts” (Mackey & Abbuhl, 2005, p. 210). Learners can use this information to modify or reformulate a non-targetlike aspect of learner output and, furthermore, restructure their interlanguage. In earlier SLA literature, researchers often subsumed feedback under the more general term ‘negative evidence.’

One of the long-standing, predominant beliefs held by both second language researchers and teachers is that feedback, as an external condition, plays an important role in the process of learning an L2. For example, White (1989, 1991) maintained that certain aspects of the L2 grammar cannot be acquired solely from the positive evidence available in the input. Carroll and Swain (1993) further argued that provision of some kind of negative feedback is necessary for L2 learners to produce hypotheses that explain available L2 data in a more reasonable manner, which in turn facilitates SLA processes.

A larger volume of the SLA research on the role of feedback has been carried out within the framework of the ‘interaction approach’ (Long, 1981; Long, 1996; Gass & Mackey, 2007) in SLA, since feedback constitutes a significant component of how interaction brings about language learning. Empirical work
conducted in this vein has investigated the relationship between oral feedback (i.e., feedback delivered in the oral mode), especially implicit negative feedback such as recasts and SLA (e.g., Braidi, 2002; Han, 2002; Iwashita, 2003; Leeman, 2003; Long, Inagaki, & Ortega, 1998; Lyster & Ranta, 1997; Mackey, 1999; Mackey, Gass, & McDonough, 2000; Mackey & Philp, 1998; Oliver, 1995, 2000; Philp, 2003; see Long, 2007, and Nicholas, Lightbown, & Spada, 2001, for reviews). Most of these works have lent support to the overall facilitative role of oral feedback in L2 development. Recent meta-analyses by Mackey and Goo (2007) and Russell and Spada (2006) have also shown a positive association between oral feedback and L2 learning.

The effectiveness of written feedback (i.e., feedback given in the written modality) has received a lot of attention from L2 writing researchers since the early 1980s. Much of the research has investigated whether written feedback provided on learners’ written compositions helps learners to improve their L2 writing accuracy (e.g., Ashwell, 2000; Chandler, 2003b; Fathman & Whalley, 1990; Ferris & Roberts, 2001; Frantzen, 1995; Hyland, 1998, 2003; Kepner, 1991; Lalande, 1982; Polio, Fleck, & Leber, 1998; Robb, Ross, & Shortreed, 1986; Semke, 1984; Sheppard, 1992). However, it has remained controversial whether written feedback actually plays a facilitative role in L2 development (e.g., Chandler, 2003; Ferris, 1999; 2004; 2007; Truscott, 1996, 2007). The empirical research base to date provides limited and inconsistent evidence to support a link between written feedback and L2 learning.
learning, as will become clear in the review of previous empirical findings presented in the following section.

The fact that written feedback is usually given in a delayed time (i.e., not immediately after learner’s erroneous output has been produced, or, not while learners are engaged in a production task) may pose a disadvantage for written feedback. Nevertheless, due to the modality in which written feedback is presented, learners can have distinctive opportunities to focus on form when receiving written feedback (Cumming, 1990; Polio et al., 1998; Qi & Lapkin, 2001; Swain, 1998). While monitoring and revising linguistic forms and structures, learners may acquire more control over their L2 knowledge and writing processes, “an opportunity that might not be as possible in speech, given its fleeting nature and the exigencies of real-time conversation” (Sachs & Polio, 2007, p. 69). Thus, it would follow that further research based on a careful review of literature is needed to uncover the effects of form-focused written feedback on L2 learning.

2.1.2. Empirical Studies on the Effects of Written Feedback on L2 Learning

2.1.2.1. Empirical Findings on Effectiveness of Written Feedback

Both L2 writing researchers and classroom practitioners have generally presumed that provision of written feedback facilitates learners’ acquisition of targeted linguistic forms and/or structures and improves the accuracy of their writing. As a result, varied methods of written feedback or error correction have been widely
incorporated into L2 writing instruction to promote written accuracy, and some L2 writing researchers have sought to gain empirical evidence to corroborate this commonly held view on written feedback. Many studies have been carried out to determine whether written feedback is effective at enhancing learners’ ability to use the L2 accurately, yielding mixed results. A close examination of the previous empirical studies reveals the need to carefully review and interpret their findings.

Among the studies that addressed the effectiveness of written feedback in L2 writing, many did not include a control group who did not receive written feedback of any kind (e.g., Ferris, 2006; Kepner, 1991; Lalande, 1982), which makes it questionable whether or not the reported improvements in written accuracy were attributed solely to the effect of written feedback provided. Some written feedback studies that involved a control group still provided content-based feedback (e.g., comments on content and/or organization) for their control group, generally due to ethical reasons (e.g., Bitchener, Young, and Cameron, 2005; Kepner, 1991). To date, six published studies have compared the effects of the presence and absence of written feedback (provided on form) on linguistic accuracy (Ashwell, 2000; Bitchener, 2008; Bitchener & Knoch, 2008; Fathman & Whalley, 1990; Ferris & Roberts, 2001; Sheen, 2007), reporting positive results for the facilitative role of written feedback.

Fathman and Whalley (1990) investigated how effective teacher feedback on form (grammatical errors) and feedback on content are in improving student writing. Seventy-two students enrolled in intermediate ESL college composition classes at
two different colleges participated in the study. They were randomly divided into four different treatment groups: (i) a grammar feedback group who were informed of the location of specific grammatical errors by means of underlining; (ii) a content feedback group who received general comments that were not text specific on the content of their compositions; (iii) a grammar and content feedback group who received a combination of grammar and content feedback; and (iv) a control group who received no feedback. Participants were asked to write a story about a sequence of eight pictures during 30 minutes. The compositions were returned to the learners a few days later with feedback which varied depending on the treatment condition. Then the participants made revisions in another 30 minute-session. Both the original compositions and the rewrites were assigned a holistic score for grammar and content respectively based on the “ESL Composition Profile” (Jacobs, Zinkgraf, Wormuth, Hartfield, & Hughey, 1981). Results indicated that learners significantly improved in grammatical accuracy in their revision when teachers provided feedback on grammatical errors. No significant differences were found between the grammar feedback group and the grammar and content feedback group. In terms of content scores, all the groups, including the control group, significantly improved regardless of the presence and type of feedback.

In another classroom study, Ashwell (2000) investigated the differential effects of order of presentation of form and content feedback. This study was set out to empirically test whether a content feedback followed by form feedback pattern of
teacher response, the generally suggested pattern of teacher feedback within a process writing approach (Zamel, 1985), is more beneficial than other patterns of feedback. He compared a content feedback followed by form feedback pattern with a form feedback followed by content feedback pattern, a pattern of mixed form and content feedback at both feedback stages, and a control condition. Fifty Japanese EFL students enrolled in two college writing classes were asked to write a total of three versions of a composition. Participants received different patterns of feedback on their first and second versions of writing during each of 12-minute feedback stages, depending on the treatment condition they were assigned to. Form feedback consisted of underlining or circling grammatical, lexical, and mechanical errors or of using cursors (∧) to indicate omission, while content feedback was mainly targeted at supra-sentential issues such as organization, paragraphing, cohesion, and relevance. In terms of the formal accuracy, results showed that there were no significant differences between the three feedback groups, and all three feedback groups improved significantly more than the control group. No significant differences were found among groups for content scores.

Beneficial effects of written feedback on writing accuracy were also documented by Ferris and Roberts (2001). They aimed to investigate how explicit error feedback needs to be to help learners self-edit their texts. In this study, 72 university ESL learners were randomly assigned into one of three feedback conditions: feedback consisting of underlining and coding all instances of errors in
five categories (verbs, noun endings, articles, word choice, and sentence structure); feedback involving errors underlined but not coded; no feedback (control).

Participants wrote a 50-minute essay, and, two weeks later, received their texts with feedback provided. They were then asked to spend 20 minutes self-editing their essays. Results showed that both groups who received feedback significantly outperformed the control group on the self-editing task. However, no statistically significant differences were found in editing success between the two groups who received different types of feedback.

The positive findings reported in these three studies, however, are vulnerable to criticism in that what they proved in the studies does not, in fact, lend support for the facilitative role of written feedback in L2 learning (e.g., Truscott, 1996, 1999, 2007). In other words, because these studies examined the effects of written feedback in terms of learners’ improvement of accuracy in revisions of their original composition, it is difficult to determine whether learning took place and learners therefore would be able to write a new piece of writing more accurately at a later time. The findings from these three studies, therefore, should only be taken as suggestive if the ultimate interest lies in elucidating the effects of written feedback on L2 development.

Unlike the three studies, Polio, Fleck, and Leder (1998) did not find that written feedback significantly improved L2 writers’ linguistic accuracy. In this longitudinal study, 65 undergraduate and graduate ESL learners enrolled in an
English for academic purposes composition course were randomly assigned to one of two treatment groups. A control group wrote four journal entries each week for seven weeks and received no feedback. An experimental group, on the other hand, wrote a journal entry followed by a grammar review and editing exercises, wrote another journal entry, and then revised one of the two journal entries. Feedback (error correction) was provided on both the editing exercises and the journal entries. For the purpose of assessment, each participant was asked to produce four essays: a 30-minute pretest essay with a 60-minute revision, and a 30-minute posttest essay with a 60-minute revision. Each essay was analyzed using two measures that were related to error-free T units. Results showed that both groups demonstrated a significant improvement on measures of linguistic accuracy both over a semester (from pre- to posttest) and from original compositions to revisions. The experimental group did not outperform the control group, however. The researchers therefore concluded that grammar correction (as practiced) was not effective.

However, it should be noted that the difference between their control group and experimental group consisted not only of inclusion of error correction, grammar reviews, and editing exercises but also of the number of writing journal entries. Had the experimental group had opportunities to write as many journal entries as their control group, different results might have been obtained. That is, the effects of the combination of error correction, grammar reviews, and editing practice might have been offset by the possible benefits which their control group might have received.

17
from engaging in a higher amount of writing process. Moreover, since error correction (written feedback) was provided together with grammar reviews and editing exercises, it is not possible to determine whether the findings can be attributed to the effect of one or more variables and, furthermore, to which variable(s).

More recently, given that one plausible explanation for the inconclusive findings of previous research may be related to the unfocused approach (i.e., treating a broad range of error categories at the same time), several researchers have carried out empirical investigations into the effectiveness of written feedback focusing on a specific linguistic target (as opposed to all types of linguistic error). In a quasi-experimental design, Sheen (2007) investigated the relative effect of two types of written feedback on ESL learners’ acquisition of articles (a as first mention and the as anaphoric reference). Ninety-one intermediate ESL learners from intact classes at an American community college were assigned to one of three treatment groups: a direct-only correction group; a direct-metalinguistic group; a control group that completed only pre- and posttests. The two experimental groups underwent two narrative task sessions during which they rewrote a story they had read and, in the following class, they were asked to look over their narratives with corrections. It was found that both experimental groups outperformed the control group on the immediate posttests (a speeded dictation test, a writing test, and an error correction test), but the direct metalinguistic group significantly outperformed the direct-only
correction group on the delayed posttests. The researcher concluded that written feedback targeting one linguistic feature was effective in improving learners’ accuracy, particularly when metalinguistic feedback is provided. However, the control group in this study did not complete the treatment tasks, which makes it difficult to determine whether the feedback was the only factor that led to the differences between the experimental and the control groups.

Similar to Sheen’s (2007) study, to address the efficacy of written feedback, Bitchener (2008) used a focused approach targeting the same linguistic target: two functional uses of the English article system (a as first mention and the as subsequent mentions). Participants were 75 international ESL learners from four intact low intermediate classes in New Zealand. Each class was randomly assigned to one of four groups: a group receiving direct error correction above each targeted error as well as written and oral meta-linguistic explanation; a group receiving direct error correction accompanied by written meta-linguistic explanation; a group receiving direct error correction only; and a control group. In this study, the control group did not receive feedback on the targeted features but, for ethical reasons, was given “a single, brief, general comment such as ‘clear outline of what is happening in the picture’ or ‘well organized piece of writing’” (p. 110). Participants produced three pieces of writing, a description of what was happening in a given picture, which served as pretest, immediate and delayed posttests. Feedback was provided on all three pieces of writing, but the researcher only provided different types of
feedback according to the treatment group assignment on the first piece of writing (pretest). Direct error corrections were only given on the second (immediate posttest) and third (delayed posttest) pieces of writing of all three treatment groups. Bitchener reported that learners who received written feedback outperformed the control group on the immediate posttest and that this level of performance was retained over a 2-month period. However, as the researcher himself recognized, the feedback group learners’ level of retention and the control group’s significant improvement from immediate to delayed posttests may have been due to additional variable from external sources. Moreover, some (if not all) participants in the control group in this study received some kind of feedback on content that may serve as an additional input on the linguistic target (i.e., article), which might have affected the results of the study.

Bitchener and Knoch (2008) extended Bitchener (2008)’s work by including more participants with different background and addressing whether migrant and international students differ from each other in the extent to which they benefit from written feedback. The researchers presented findings from the final data set which consisted of 69 migrants who had settled in New Zealand within 18 months and were enrolled in a university as well as from 75 international ESL learners in the same country. Four intact low intermediate classes from each of three institutions (n = 144) were randomly assigned to one of four treatment conditions, varying type of feedback (direct error correction, written and oral meta-linguistic explanation; direct
error correction and written meta-linguistic explanation; direct error correction only; control). As is the case with Bitchener (2008), it was only on the first piece of writing (pretest) that a researcher provided different types of feedback depending on the treatment group assignment. On the second and third pieces of writing (posttests), a tick was put above correct uses of the targeted features and a cross above incorrect uses, but no direct feedback (i.e., correct forms) was provided. It was reported that learners who received three types of feedback outperformed those who did not, and that their level of accuracy was retained over a 7-week period. Bitchener and Knoch (2008) also reported no difference in the extent to which migrant and international learners improved their writing accuracy as a result of written feedback.

To summarize then, although some researchers have sought to empirically investigate the effects of written feedback targeting linguistic forms or structures on the development of linguistic accuracy in L2 writing, the research base currently does not provide any clear-cut and conclusive answers. A limited number of studies have compared written feedback on form versus the lack thereof and the subsequent effect of written feedback on learners’ L2 development measured through new pieces of writing. Only three empirical studies (Bitchener, 2008, Bitchener & Knoch, 2008; Sheen, 2007) to date have reported positive evidence for the facilitative role of written feedback while one study (Polio et al., 1998) revealed no significant effect. As mentioned earlier, the findings from these studies still need to be taken with care. The most plausible explanation to account for the mixed results appears to relate to
design and methodological issues, which are addressed in a later section. More carefully designed studies on the effects of written feedback on L2 written accuracy, therefore, are shown to be warranted. At the same time, although definitive conclusions have yet to be reached as to the effectiveness of written feedback, a great deal of empirical research has been carried out to identify the optimal means of providing written feedback in order to promote L2 learning. The next section reviews the studies addressing relative effectiveness of different types of written feedback.

2.1.2.2. Empirical Findings on Relative Effectiveness of Different Types of Written Feedback

The core assumption in the use of written feedback is that it directs learners’ attention to formal aspects of the L2, thereby facilitating L2 development and, at the same time, leading to greater accuracy of learners’ subsequent production. Based on this idea, a group of studies has investigated the extent to which learners benefit from different types of written feedback or combinations of them in L2 writing. These studies have employed various pedagogical techniques to draw learners’ attention to their grammatical deficiencies. Although written feedback is mostly provided on learners’ written production as to grammatical accuracy, studies sometimes have included comments and questions on the content of learner compositions and compared form-based feedback to content-based feedback (e.g., Semke, 1984; Sheppard, 1992). The techniques used to provide form-based feedback
include many types, from the most indirect or implicit method such as marginal feedback to the most direct or explicit explanation of the rules, metalinguistic explanation.

One classification which has widely been adopted in many empirical written feedback studies is direct versus indirect feedback. Direct feedback refers to a process in which learner errors are corrected while indirect feedback is when the existence of errors are indicated in some way, but errors are not corrected. Direct feedback, in general, is given by means of the provision of the correct linguistic form or structure above or near linguistic error. Indirect feedback is provided through underlining, circling, or highlighting errors, marking the number of errors made per line in the margin, or using a code to indicate the location and type of error. Learners are therefore required to engage in problem solving to correct an error, when provided with indirect feedback.

2.1.2.2.1. Empirical findings on relative effectiveness of direct and indirect feedback.

The relative efficacy of direct and indirect feedback is of primary interest in the body of literature. Lalande (1982), for example, examined the relative effect of direct and indirect feedback on the grammatical and orthographic accuracy of written compositions. Sixty learners enrolled in four classes of intermediate-level German at an American university were selected for this study based on their grade point averages (GPAs) and their first essay (pretest) score. All participants wrote
five in-class essays with a 250 word-length (60 minutes for essays 1 and 5, and 45 minutes for essays 2, 3 and 4), and completed three rewrite tasks (50 minutes each). The first and last essays served as pre- and posttests. Two classes, which made up the experimental group, received coded feedback (i.e., indirect feedback) by means of a systematic marking on their essays. They were required to interpret the given correction codes, correct their mistakes, and then rewrite the essay. The two other classes, which formed a control group, received error corrections (i.e., direct feedback) from their teachers and were asked to incorporate them into a rewritten version. It was found that the experimental group receiving coded feedback showed a significantly better performance on the posttest than did the direct feedback group. The researcher concluded that the coded feedback (indirect feedback) consisting of error awareness and problem-solving techniques had a significant positive effect on the development of learners’ writing skills.

Semke (1984) conducted a similar study to Lalande (1982) which yielded different results. She compared the effects of four types of feedback on free-writing assignments (journal entries) over a 10-week period. Participants were 141 learners enrolled in eight sections of first-year German at an American university. Each two sections received different types of feedback: (1) writing comments and questions (on content) rather than corrections; (2) marking all errors and corrections; (3) a combination of positive comments (on content) and corrections; and (4) indicating errors by means of a code and requiring learners to make their own corrections and
rewrite the assignments. All participants completed a pre- and a posttest consisting of a timed, free-writing sample and a timed, multiple-choice cloze test. No significant differences were found among the four treatment conditions. In addition, results indicated that only those who received feedback type (1) increased in fluency and general language proficiency. The researcher concluded that learners’ L2 writing ability was enhanced by practice alone, and error corrections did not increase writing accuracy, fluency or general language proficiency. However, these findings need to be taken with caution due to some limitations in research design. The type of feedback was not the only dissimilarity between the experimental groups in this study: the groups also differed in terms of the length of writing they were required to produce. The group who received feedback type (1) had to write as twice long as those who received feedback type (2) and (3), and the group who received feedback type (4) wrote only about one-half as much as new material because of the time spent to make revisions. In addition, the dependent measure of written accuracy (a 10-minute free-writing activity) and the instruction given to participants seem to be somewhat biased in favor of the group that did not receive corrections.

Robb, Ross, and Shortreed (1986) investigated the effects of different types of written feedback which differed in terms of the degree of salience on improving learners’ overall L2 writing quality over one academic year (from mid-April until mid-January). Participants were 134 EFL learners enrolled in four sections of an English composition course at a Japanese university. Four different methods of
providing feedback were compared: (1) direct correction, indicating the errors and
providing correct forms; (2) coded feedback, indicating location and type of error by
means of an abbreviated code system; (3) uncoded feedback, highlighting errors
with a yellow text-marking pen to indicate the location of errors; and (4) marginal
feedback, writing the total number of errors made per line in the margins of the
student’s paper. They were asked to revise their weekly essay assignments based on
the feedback they received, and wrote five narrative compositions which served as
tests at equal intervals during the academic year. Results showed that all groups
improved their writing accuracy over time and no significant differences were found
among the different treatment groups. Robb et al. concluded that the more direct
methods of feedback do not tend to yield results commensurate with the instructor’s
time and effort, and that less time-consuming methods of drawing learners’ attention
to error may suffice.

Contrary to the findings of previous studies, Chandler (2003) found a
significant advantage for direct feedback in her investigation into the differential
effects of direct and indirect feedback in two studies. In the first study reported in
Chandler (2003), 31 high intermediate/advanced ESL learners in two sections of a
writing course wrote five essays as homework assignments over a semester. The
experimental group was required to revise each assignment, correcting all the errors
underlined by the teacher-researcher, whereas the control group who received
indirect feedback in the form of underlining was not asked to revise. The
The experimental group received direct feedback for any remaining errors on the rewrites. Results showed that the experimental group demonstrated a significant improvement from the first to the fifth pieces of writing while the control group did not. A statistically significant difference was revealed in improvement of accuracy between the two groups over the 10 weeks. It is not possible, however, to determine whether the observed difference between the two groups can be attributed solely to the different type of feedback or whether it was actually caused by the combined effects of feedback type and presence/absence of rewriting.

2.1.2.2. Empirical findings on relative effectiveness of different types of indirect feedback.

In comparing the effects of different written feedback varying in the directness/explicitness on L2 writing accuracy, very few studies have examined the relative merits of different types of indirect feedback (Robb et al., 1986; Ferris & Roberts, 2001). Robb et al. (1986) employed different methods of indirect feedback in their study, using coded, uncoded and marginal feedback. Ferris and Roberts (2001) also used coded and uncoded feedback in their experimental design, and compared their relative effectiveness. Statistically significant differences were not found between coded and uncoded feedback in either of these studies. One point worthy of attention here is that Ferris and Roberts (2001) measured the effect of feedback on learners’ linguistic accuracy through text revisions, not through subsequent new pieces of writing. As discussed above, therefore, the finding from
this study should be taken as only suggestive. Regardless, it is not possible to draw any conclusion as to the relative effectiveness of different types of indirect feedback based on the very limited research available.

2.1.2.2.3. **Empirical findings on relative effectiveness of different types of direct feedback.**

In recent years, the efficacy of different types of direct feedback has received more attention from researchers than that of different types of indirect feedback. Several studies by Bitchener and his colleagues and Sheen (Bitchener et al., 2005; Sheen, 2007; Bitchener, 2008, Bitchener & Knoch, 2008) have investigated the relative effectiveness of different techniques of providing direct feedback.

Bitchener et al. (2005) addressed whether two types of direct written feedback affect learners’ ability to accurately use three linguistic forms (prepositions, the past simple tense, and the definite article) in new pieces of writing. Fifty-three adult migrant learners enrolled in a post-intermediate ESOL program in New Zealand were assigned to three treatment groups. Participants completed four writing tasks over a 12 week period in weeks 2, 4, 8, and 12. Nineteen learners enrolled in a class meeting for 20 hours per week received direct written feedback consisting of full, explicit corrections above the underlined errors together with a 5 minute student-researcher oral conference. Seventeen learners in a class meeting for 10 hours per week received direct feedback only. The other seventeen learners enrolled in a 4 hours per week class received no feedback on the targeted forms, but
on ethical grounds were given feedback on the quality and organization of their content. Using two-way repeated-measures ANOVAs, the researchers found a significant effect for the combination of direct feedback and conference in terms of the accurate use of the past simple tense and the definite article. Based on this finding, they concluded that direct written feedback in combination with direct oral feedback had a greater effect than direct feedback on improving learners’ accuracy over time when the simple past tense and the definite article were used on new pieces of writing.

However, some caution is called for in the interpretation of data from this study. Given that the significant main effect for feedback type indicates that the group receiving direct feedback accompanied by conference performed differently from the other groups regardless of time, the researchers’ conclusion is difficult to reconcile with the actual results of ANOVA analysis. In order to conclude that the combined feedback was more effective than direct feedback in improving learner’s accuracy over time, it would have been necessary to find a significant interaction effect suggesting that the group’s pattern of change from weeks 2 to 12 had differed from the other group’s trajectory. A significant interaction effect of feedback type and time was reported for the combined feedback when the prepositions were targeted, whereas the same effect was not observed in the cases of the simple past tense. This can also be confirmed by comparing the changes in the actual mean scores for three groups presented in their table. Moreover, the fact that the treatment
group was assigned depending on the amount of L2 instruction that the participants received per week limits both the internal and external validity of the results. In other words, it is difficult to rule out the possibility that the between-group difference in amount of instruction might have affected the results, and the non-random group assignment limits the generalizability of the results beyond the sample. Nonetheless, as Bitchener et al. asserted, the results of this study could be taken to indicate that the effect of the direct written feedback-conference condition differed depending on the linguistic targets corrected within the context of this study.

Sheen (2007) compared the differential effect of direct-only correction and direct metalinguistic correction on intermediate ESL learners’ article acquisition. Direct-only correction was comprised of indicating the location of an error and providing a correct form, while direct metalinguistic correction consisted of indicating the location of an error, providing a correct form, and supplying metalinguistic comments that explain the correct form. She found no significant difference between the two types of direct feedback on the immediate posttests (a speeded dictation test, a writing test, and an error correction test), but she did find that direct metalinguistic correction was more effective than direct-only correction on the delayed posttests.

In another study targeting two functional uses of the English article system (indefinite *a* for first mention and definite *the* for subsequent mentions), Bitchener (2008) compared three types of direct feedback: (1) direct error correction above
each targeted error as well as written and oral meta-linguistic explanation; (2) direct error correction with written meta-linguistic explanation; and (3) direct error correction only. In this study, the written metalinguistic explanation contained an explanation of the two targeted functional uses of the articles and an example of their use. Oral metalinguistic explanation took the form of a 30-minute mini-lesson during which learners received an explanation of the rules and examples and completed a short grammar exercise. Seventy-five low intermediate ESL learners wrote three compositions (pre-, immediate post-, and delayed posttests), and received different types of feedback on their first composition only. As pointed out in the previous section, all three groups in this study received direct feedback on their second and third compositions, regardless of the treatment group assignment. A direct and valid comparison of three types of direct feedback, therefore, can be made only in terms of gain scores between pre- and immediate posttests. However, it was not reported whether there was any significant difference among the three groups’ gain scores, although all three groups made a significant improvement from pre- to immediate posttests. Instead, comparing the mean scores of each group regardless of time, the researcher reported that two groups who received feedback type (1) and (3) outperformed the control group who did not receive feedback on the targeted forms (but received a brief, general comment) while the group receiving feedback type (2) did not.

Extending Bitchener (2008)’s study, Bitchener and Knoch (2008) included
69 migrant learners as well as 75 international ESL learners. Participants received
the same three types of direct feedback on their first composition (pretest) as in
Bitchener (2008). According to the study report, a tick was put above correct uses of
the targeted features and a cross above incorrect uses on the second and third writing
texts (posttests), but no direct feedback (i.e., correct forms) was provided. Given this,
in order to accurately compare the effectiveness of three types of direct feedback, it
would have been necessary to test whether there was any significant difference
among the three groups’ trajectories from pre- to immediate posttests. This, however,
was not directly addressed in the study. The researchers reported that none of the
three direct feedback groups differed from each other based on the comparison of
mean accuracy scores between groups (regardless of time). The finding that all three
direct feedback groups made a significant improvement from pre- to immediate
posttests was consistent with the finding of Bitchener (2008).

To sum up, a good deal of research has attempted to determine what type of
written feedback provided on L2 learners’ written texts better facilitates their
linguistic accuracy, with inconsistent results. One earlier study (Lalande, 1982)
provided supporting evidence for indirect feedback, while two studies (Robb et al.,
1986; Semke, 1984) reported no significant difference between indirect and direct
feedback, and one study (Chandler, 2003) reported positive evidence for direct
feedback. One study (Robb et al., 1986) that also addressed the effects of different
types of indirect feedback revealed no significant difference between coded and
uncoded feedback. Recent studies have investigated the relative merits of different kinds of direct feedback using a focused approach. One study (Bitchener et al., 2005) suggested that the effects of direct feedback with an oral conference may vary depending on the targeted linguistic form. A more recent study (Sheen, 2007) yielded supporting evidence that more direct feedback (direct metalinguistic correction) is beneficial in the long-term. Other studies found positive effects for all three types of direct feedback on learners’ improvement from pre- to immediate posttests, but the relative efficacy of three direct feedback types remains inconclusive (Bitchener, 2008; Bitchener & Knoch, 2008). Therefore further research is clearly warranted in order to isolate which feedback type is most effective. Some issues that need to be addressed in further research and which emerged from limitations of previous studies are discussed in next section.

2.1.3. Remaining Issues

2.1.3.1. Research Design and Methodological Issues

As alluded to earlier, various differences present in the body of written feedback research make it difficult to directly compare these studies. Nevertheless, a close examination of several factors seems to be in order to help identify the directions for future research on the role of written feedback in SLA.

2.1.3.1.1. Measurement of learning.

First, as mentioned above, many of the previous written feedback studies
have explored the effects of written feedback or error corrections on learners’ subsequent performance in revision tasks of the same essay after having received written feedback on it (e.g., Ashwell, 2000; Fathman & Whalley, 1990; Ferris & Roberts, 2001). It should be noted that revision studies cannot explain or predict anything about the effects of written feedback on interlanguage development or learning outcomes, although the findings from research of that kind can be taken as suggestive. More research employing a pre-test/treatment/post-test design needs to be conducted to investigate whether written feedback on learners’ written compositions indeed promote language learning.

2.1.3.1.2. Control group.

Another point to take into account in designing a study on the effectiveness of written feedback is the inclusion of a control group. Many of the previous studies of written feedback did not compare the effects of its presence and absence, but instead compared the effects of different types of feedback (e.g., Kepner, 1991; Lalande, 1982; Robb et al., 1986; Semke, 1984). Even in cases where control group learners did not receive feedback on linguistic errors, they were provided with feedback on content or organization, which makes it hard to document the impact of written feedback on linguistic accuracy. In relation to this, Ferris (2004) stated that this may be because most teachers find it unethical not to provide feedback to students who usually expect it. Truscott (2007) argued that provision of comments on content and clarity seems to be generally accepted and went so far to say that no
one recommends comparing the effects of correcting errors with providing no feedback at all (p. 4).

From a research perspective, however, the lack of a control group constitutes a major shortcoming which cannot be justified on ethical or conventional grounds and limits the internal validity of these studies. The optimal research design would be one in which an experimental group that receives written feedback on linguistic forms is contrasted with a control group that receives neither feedback on the forms being targeted in the experimental condition nor any other feedback on content on dependent measures of linguistic accuracy. Data from this kind of design will form the basis for an accurate analysis for the effectiveness of written feedback versus no written feedback in improving L2 learners’ linguistic accuracy.

2.1.3.1.3. Confounding variables.

There are also potential confounding variables in several of the written feedback research. Studies that included revision opportunities after provision of feedback, error logs, supplemental grammar instruction, or student-teacher conferences as part of the treatment condition preclude us from determining whether the differential performances between the treatment and control (or comparison) groups were caused by the feedback or by other confounding variables (e.g., Polio et al., 1998; Sheppard, 1992). Studies that measured learners’ improvement in homework essay assignments (e.g., Chandler, 2003) weaken the findings because it is not clear whether differences found between groups resulted from the given
feedback or the aid of external resources (e.g., textbooks, dictionaries). Likewise, many of the studies did not report the amount of time spent on the dependent variable tasks. This lack of information makes it difficult to exclude a possibility that the effects of time-on-task could have influenced the results. Thus, care needs to be taken to control any potential external variables in designing and conducting future research in this area.

2.1.3.2. Possible Role of Type of Linguistic Target in Mediating Written Feedback and L2 Development

Mackey (2006), in a study which investigated the relationship between L2 learners’ noticing of feedback and their subsequent L2 development, found that the degree of noticing reported by learners varied depending on the type of linguistic item within the linguistic domain of morphosyntax. Learners reported the noticing of English question forms more often than past tense or plurals. The results also suggested that noticing and L2 learning may be connected in terms of development of question formation. These findings were discussed in the context of the different degrees of communicative value of the linguistic targets.

In the strand of Computer-Assisted Language Learning, a study reported in Nagata (1993) and Nagata and Swisher (1995) compared the effectiveness of two types of computer-delivered feedback on L2 learners’ written production of Japanese passive structures. Thirty-two university students enrolled in a second-year Japanese course were randomly assigned into either a traditional or an intelligent
feedback condition. While completing a series of four computerized lessons on passive structures in Japanese, the traditional group received feedback only indicating what grammatical information was missing or incorrect in their responses by a traditional computer-assisted language instruction (T-CALI) system, whereas the intelligent group received identical feedback in addition to detailed metalinguistic explanations about their errors by an ‘intelligent’ version of the system (I-CALI). Quantitative analyses revealed that the I-CALI group significantly outperformed the T-CALI group on production of particles, although type of feedback did not have any significant effect on learners’ production of verbal predicates. Learner responses on a questionnaire administered following the treatment suggested that students overall preferred intelligent feedback to traditional feedback. These findings led Nagata to conclude that intelligent feedback (metalinguistic feedback) is more effective than traditional feedback when the targeted grammatical system of the L2 is nontrivially complex.

Some studies in the textual input enhancement strand of research have investigated type of linguistic item as a separate independent variable. Leow et al. (2003), for example, found that learners exposed to perceptually more salient forms (Spanish present perfect) reported significantly more noticing than those exposed to less salient forms (Spanish present subjunctive), although the aforementioned effects for type of linguistic item were not found for reading comprehension or intake of the target forms.
Taken together, these findings seem to concur with the hypothesis that different linguistic forms are not processed in the same way (VanPatten, 1994). Given the support for this idea, the effectiveness of written feedback may depend on the type of linguistic target. In addition, it seems interesting to further explore the relative efficacy of type of written feedback in relation to the type of linguistic target.

2.1.3.3. Learner-internal Factor Linking Written Feedback and L2 Learning

As mentioned above, the rationale behind research on written feedback seems to be drawn from theoretical models of attention and awareness in SLA, which are discussed in detail in section 2.2. In empirical studies of written feedback, a major assumption is that written feedback provided on learners’ written production would draw learners’ attention to the linguistic features marked/corrected, thereby promoting superior, subsequent awareness and further processing of the attended forms, and affect learning when compared to no feedback condition. Given these premises regarding attention and awareness, it is crucial to operationalize and measure them directly.

However, written feedback studies to date have failed to use data elicitation procedures to confirm that learners did indeed pay attention to the provided feedback and that the given feedback did actually lead to superior, subsequent awareness and further processing of the attended linguistic targets. They only made an indirect inference regarding the presence and the focus of attention and awareness as well as learners’ subsequent processing of L2 data by means of post-exposure
tests. It is essential to measure attentional processing and to gather more direct evidence through on-line (during exposure) measures in order to guarantee the internal validity of research premised on attention and awareness in SLA (Leow, 1999b).

From both a theoretical and a practical point of view, it is of great interest and significance to investigate whether the provision of written feedback and, furthermore the type of written feedback would affect L2 learners’ attentional processing of targeted linguistic structures, which may, in turn, promote their learning. Oral interactional feedback suggests that the effect of feedback on subsequent L2 learning may depend on whether learners notice its corrective function, how they perceive its focus, and/or what has been targeted in feedback (e.g., Mackey, 2006, Mackey et al., 2000). This indicates a relationship between the instructional benefits of feedback and learners’ noticing (attention and a low level of awareness). Moreover, with respect to task design and grammar pedagogy, Loschky and Bley-Vroman (1993) implies that different types of feedback lead learners to engage in different levels of cognitive processes by triggering different levels of awareness. Thus, further research is required to corroborate the relationships of these factors, such as written feedback, learners’ awareness, and learning, and also to elucidate a possible relationship between type of written feedback and levels of awareness. It also seems meaningful to explore whether learners’ attentional processes are differentially influenced by the type of linguistic structures targeted in
feedback to better understand the effect of written feedback on L2 learning. Section 2.2 further discusses the issues of cognitive processes involved in SLA.

2.2. Attention, Awareness, and SLA

2.2.1. Theoretical Models of Attention and Awareness in SLA

Acknowledging that not all input to which L2 learners are exposed is taken in and integrated into learners’ interlanguage systems, SLA research over the last decades has begun to investigate more precisely how L2 learners process input. Drawing on the most accepted position, supported by empirical findings, on the role of attention in general learning in the field of cognitive psychology (e.g., Carlson & Dulany, 1985; Carr & Curran, 1992; Nissen & Bullemer, 1987; Posner, 1992; Reber, 1967, 1976, 1989, 1993), SLA researchers have inquired into the role that attention and awareness play in driving learners to acquire an L2 (e.g., Tomlin & Villa, 1994; Schmidt, 1990, 1993, 1994, 1995, 2001; Robinson, 1995). The most widely accepted position in cognitive psychology states that attention to stimuli is necessary for long-term memory storage and that little, if any, learning can occur in the absence of attention.

SLA literature has witnessed a lively discussion and debate over the roles of attention and awareness in recent years. As pointed out by Schmidt (1995, 2001), it is a thorny question to clearly differentiate between attention and awareness, largely because of the general assumption in psychology that the two constructs are two
sides of the same coin (James, 1890; Posner, 1994). Attention is, according to Schmidt (1995), “one of the basic mechanisms in an information processing system” (p. 18). Attention also refers to “a limited-capacity system,” “the process of selecting critical information for further processing.” “Effortful processing that can be contrasted with more automatic and less effortful processing” and “a matter of the control of information and action” (Tomlin & Villa, 1994, p.187; Robinson, 1995). Schmidt (1995) distinguishes awareness from attention by defining awareness as the correlated subjective experience of “noticing what one attends to, remarking upon it, being aware of it” (p. 18). Based on SLA (e.g., Schmidt, 1990) and cognitive science (e.g., Schacter, 1992), Tomlin and Villa (1994) defined, in a similar manner, awareness as “a particular state of mind in which an individual has undergone a specific subjective experience of some cognitive content or external stimulus” (p. 193).

awareness for subsequent processing of input to take place). Some SLA researchers have proposed that it is possible to learn without awareness (e.g., Tomlin & Villa, 1994), whereas others have posited that awareness is a necessary condition for L2 learning (e.g., Robinson, 1995, 1996, 2003; Schmidt, 1990, 1993, 1994a, 1994b, 1995).

Based on the previous work by Posner (1992), Tomlin and Villa (1994) proposed a finer-grained analysis of attention for SLA research. In their functional model of input processing, which assumes that attention is a limited-capacity system, they identified three mechanisms or subsystems of attention: (i) alertness, “general readiness to deal with incoming stimuli or data” (p. 190); (ii) orientation, “the specific aligning of attention on a stimulus” (p. 191); and (iii) detection, “the cognitive registration of sensory stimuli” (p. 192). They argue that it is detection that is necessary and sufficient condition for further learning processes and subsequent acquisition among the three attentional processes, because “detection is the process by which particular exemplars are registered in memory and therefore could be made accessible to whatever the key processes are for learning” (pp.192-193). Alertness, which is related to motivation, and orientation, which is related to instructional methods such as input-flooding, may enhance the chances of occurrence of detection, but they are not necessarily needed. It is noteworthy that detection, in this model, does not require awareness and, thus, they claim a dissociation between awareness and learning.
In opposition to Tomlin and Villa’s (1994) perspective that learning can take place even in the absence of awareness, the most commonly accepted and influential view in the current SLA literature is Schmidt’s noticing hypothesis (1990, 1993, 1994a, 1994b, 1995, 2001), which argues for an important role of awareness in learning. According to his noticing hypothesis (1990 and elsewhere), attention controls access to awareness and is responsible for the subjective experience of noticing, which he maintains is “the necessary and sufficient condition for the conversion of input to intake” (p. 209). In other words, L2 learners must first “notice” – show focal attention and awareness of particular aspects of input – so that any further processing, beyond short-term memory, of the noticed aspects of input can take place. Schmidt considers focal attention as isomorphic with awareness and, thus, rejects any dissociation of awareness and learning. He further differentiates between two levels of awareness: awareness at the level of noticing (i.e., conscious registration of the linguistic form or elements of the surface structure in the input), which is a low level of awareness, and awareness at the level of understanding (i.e., knowing any abstract, general rules or principles of the form or surface level linguistic phenomena), a higher level of awareness (Schmidt, 1995, 2001). The latter is related to the ability to analyze, compare and test hypotheses about the linguistic input (1990), which he argues leads to restructuring and system learning while the former results in intake and item learning (1993, 1995). Schmidt proposes it is awareness at the level of noticing that is necessary for intake whereas awareness at
the level of understanding can facilitate learning, but it is not necessary.

Robinson (1995, 1996) proposes a model of attention, awareness and detection that reconciles the two perspectives of Tomlin and Villa (1994) and Schmidt (1990 and elsewhere) on the role of awareness in L2 learning. In this model, noticing is defined as “detection plus rehearsal in short-term memory, prior to encoding in long-term memory” (Robinson, 1995, p. 296). Detection, according to Robinson, is responsible for encoding in short-term memory, and is considered the first step in the learning process, prior to noticing (1996, p. 59). He views simple detection (without awareness) an insufficient condition for learning to take place. For Robinson, noticing involves awareness and is crucial for learning, which is in line with Schmidt’s view. He also supports Tomlin and Villa (1994)’s concept of detection, but assigns it to an early stage and, consequently, a less important role in language learning.

2.2.2. Empirical Studies on the Role of Awareness in SLA

Several SLA researchers have directly investigated, within an attentional framework, the role of awareness in L2 development mainly in a formal classroom context (e.g., Leow, 1997, 2000, 2001a; Martínez-Fernández, 2008; Rosa & Leow, 2004a, 2004b; Rosa & O’Neill, 1999; Sachs & Suh, 2007). Addressing the role of awareness in relation to Schmidt’s noticing hypothesis, Leow (1997) explored whether and how different levels of awareness of morphological forms in a problem-
solving task affected learners’ recognition and written production of those forms. The concurrent verbal protocols produced by 28 beginning adult learners during exposure to Spanish stem-changing past tense verbs were analyzed quantitative and qualitatively. The results indicated that participants demonstrating a higher level of awareness performed significantly better than those with a lower level of awareness (i.e., those who showed they simply attended and noticed) on both the recognition and written production post-exposure tests. It was also revealed that meta-awareness, that is, a report of being aware of the experience of some cognitive content or external stimulus, seems to correlate with an increased use of hypothesis testing and morphological rule formation while absence of meta-awareness appears to correlate with an absence of such conceptually-driven processing. The findings provide empirical evidence for the facilitative role of awareness in L2 learning supported by Schmidt (1990 and elsewhere) and Robinson (1995).

Leow and his colleagues (Rosa & Leow, 2004b; Rosa & O’Neil, 1999) extended this line of research by investigating the role of awareness at a syntactic level, and obtained the comparable results to those of Leow (1997). In these studies, Spanish contrary-to-fact conditional sentences were presented to L2 learners in five different conditions that vary in the degree of explicitness in a puzzle-solving task. Concurrent verbal reports were used to corroborate different levels of awareness. Intake was measured through learners’ performance on a multiple-choice recognition task. Rosa and Leow (2004a), a follow-up study on Rosa and O’Neill
made some changes in the research design, such as use of a computerized task, inclusion of novel exemplars, a written production post-exposure task and delayed posttests. The studies revealed that awareness at the levels of noticing and understanding contributed to a significant increase in learners’ ability to take in the targeted structure and to produce the targeted structure, including new exemplars. It was also found that the higher the level of awareness demonstrated during the experimental exposure, the stronger the effect on learners’ performance on posttests.

The only exception to these findings in this strand of SLA literature to date is Williams’ (2005) study. Extending Williams (2004), this study investigated the learning of form-meaning connections under conditions in which the forms were noticed but the crucial aspects of meaning were not. The overall results of the two experiments presented in the study showed that the learners who were classified as unaware of the unidentified, underlying rule behind the choice of determiner in noun phrases selected the correct determiner-noun combination at significantly above-chance levels for generalization items. It was also revealed that the aware learners performed significantly better than those who were not aware on the generalization test, and learners’ prior knowledge of gender languages had some effect. Williams took these findings as supportive evidence for learning form-meaning connections in the absence of awareness, and therefore, for Tomlin & Villa (1994)’s contention on the dissociation of awareness and learning. However, the oral interviews in William’s study did not demonstrate absence of awareness. What was measured and
coded as ‘aware’ in the study was, in fact, limited to awareness at the level of understanding, excluding awareness at the level of noticing, a lower level of awareness. Thus, Tomlin and Villa’s (1994) perspective that learning can take place without awareness still awaits empirical support.

In sum, the overall findings of the studies provide empirical support for the crucial role of awareness in subsequent processing of L2 data. Furthermore, higher level of awareness appears to correlate with consequent higher recognition and written production. I will now turn to the measurements of the constructs, attention and awareness, in the context of SLA.

2.2.3. Operationalization and Measures of Attention and Awareness in SLA

While a large number of SLA studies, implicitly or explicitly premised on attention, have attempted to draw learners’ attention to the targeted input, they have failed to either operationalize or establish methodologically the presence of this process in their research designs. Instead, they relied on posttest scores to infer what learners paid attention to while interacting with the input, and took subsequent language changes as evidence of attention in a traditional pretest-treatment-posttest design. The underlying assumption was that learners’ improved performance on posttests was due to the fact that the given treatment had indeed directed their attention to the target form. However, as identified by Leow (1999b and elsewhere), the lack of direct measure of attentional processes is maybe “the most important
shortcoming in measurement” (p. 65) of most SLA studies premised on attention during the 1990s.

More recently, several researchers have used various measures of attention, noticing, and awareness in their empirical investigations. The introspective measures of attention/noticing can be generally classified into on-line and off-line methods based upon the timing of collecting data on participants’ cognitive processes in specific tasks. On-line data are elicited during the actual process of performing a task while off-line data are gathered after completion of the task.

On-line measures of attention include on-line uptake charts (e.g., Mackey, McDonough, Fujii, & Tatsumi, 2001), underlining, circling, or checking targeted linguistic structures in written text (e.g., Fotos, 1993; Greenslade, Bouden, & Sanz, 1999; Izumi &Bigelow, 2000; Izumi, Bigelow, Fujiwara, & Farnow, 1999), note taking (e.g., Izumi, 2002), recording processing time (e.g., Collentine, 1997), on-line learning journals (e.g., Mackey, 2006), and on-line think-aloud protocols (e.g., Alanen, 1995; Bowles, 2004, 2008; Bowles & Leow, 2005; Leow, 1997, 1998a, 1998b, 2000, 2001a, 2001b; Leow et al., 2003; Leow & Morgan-Short, 2004; Martínez-Fernández, 2008; Qi & Lapkin, 2001; Rosa & Leow, 2004a, 2004b; Rosa & O’Neil, 1999; Sachs & Polio, 2007; Sachs & Suh, 2007; Swain & Lapkin, 2002).

Off-line methods of measuring attention include off-line questionnaires (e.g., Alanen, 1995; Izumi, 2002; Mackey et al., 2002; Robinson, 1995, 1996b, 1997a, 1997b), stimulated recall (e.g., Adams, 2003; Mackey, 2002, 2006; Mackey et al.,
2000; Mackey et al., 2002; Nabei & Swain, 2002; Swain & Lapkin, 2002), off-line uptake recall charts (e.g., Slimani, 1989), free recall (e.g., Greenslade et al., 1999; VanPatten, 1990), interviews (e.g., Izumi et al., 1999; Williams, 2005), learning diaries (e.g., Altman, 1997; Grabe & Stoller, 1997; Schmidt & Frota, 1986; Warden, Lapkin, Swain, & Hart, 1995), meta-talk (e.g., Swain & Lapkin, 1998, 2001), immediate reports (e.g., Philp, 2003), and immediate retrospective verbal reports (e.g., Egi, 2007).

Some drawbacks have been proposed concerning the use of off-line methods of measuring attention, noticing, and awareness. These methods are fallible due to memory constraints (Ericsson & Simon, 1984) and susceptible to potential reconstructive processes (Nisbett & Wilson, 1977). In addition, it is hard to ascertain information heeded during task performance from other related information gained before or after task completion. According to psychology research (e.g., Baddeley, 1990), information may be retained in short-term memory for at most as long as two seconds. Therefore, learners may no longer remember what was actually attended to after the two seconds, which may, in turn, lead to alteration in memories. That is, learners’ off-line reports may not reflect an accurate representation of what they paid attention to at the time of processing. In relation to this, Schmidt also states that “recall is also good evidence that something was noticed, but only if prior knowledge and guessing can be controlled” (p. 20). Thus, some caution is in order in using off-line methods.
Such potential shortcomings notwithstanding, Ericsson and Simon state that “retrospective reports of specific cognitive processes--provide powerful means for gaining information about such processes” (p. 30). Based on this, the present study employs post-exposure retrospective (off-line) questionnaires as one of two methods of measuring the presence and levels of awareness raised during feedback processing.

The on-line methods mentioned above are not without limitations. For example, processing time measures and behavioral measures of attention such as underlining, circling, and checking targeted linguistic structures in written text or on a piece of paper at every instance do not demonstrate to what extent such items are attended and whether such items are actually processed in learners’ mind. Additionally, the behavioral measures of attention pose risks of incompleteness (i.e., the measures may not include all the items that are attended to) and inaccurateness (i.e., the measures may not exclude all the items that are not attended to). Likewise, note taking may not provide a comprehensive and precise report of what was attended due to time constraints in task performance and the time-consuming nature of writing. On-line think-aloud protocols also have limitations, which are discussed in a later section. Nevertheless, as Schmidt (2001) who draws on Merikle & Cheesman (1987) states, “[t]he clearest evidence that something has exceeded the subjective threshold and been consciously perceived or noticed is a concurrent verbal report, since nothing can be verbally reported other than the current contents
of awareness” (p. 20). The next section presents a description of verbal reports with a focus on concurrent verbal reports, or think-aloud protocols, and the validity issues of concurrent verbal protocols.

2.2.3.1. Concurrent Verbal Reports and Reactivity Issue

Rooted in the fields of psychology and philosophy, verbal reports refer to an introspective method used to obtain data about the thought processes involved in performing a task. According to Ericsson and Simon’s (1984, 1993) seminal review of verbal protocols, verbal reports are categorized as either concurrent (on-line) or retrospective (off-line) reports based upon the timing of collecting verbalizations. Concurrent reports are gathered when participants verbalize their thoughts while performing a task, whereas retrospective reports are elicited after a task has been completed. Ericsson and Simon (1984) viewed the two types of verbal reports as “the closest reflection of cognitive processes” (p. 16). Furthermore, Ericsson and Simon (1993) went on to propose that concurrent verbal protocols should be collected whenever possible so that the actual cognitive processing and reporting occur simultaneously and that the problem of double access to information can be avoided (p. xiii).

As concurrent verbal protocols provide a direct way of accessing learners’ internal cognitive processes which otherwise would not be available, they have been extensively used in various strands of second language research. Included are strands such as learning strategies (e.g., Carrell, 1989; Davis & Bistodeau, 1993;
Warren, 1996), reading (e.g., Aweiss, 1993; Brown, 1996; Chun, 2001; Dewey, 2004; Nassaji, 2004, Wang, 2001) and writing (e.g., Johnson, 1992; Mahfoudhi, 2003; McDonough & McDonough, 2001; Qi & Lapkin, 2001; Raimes, 1987; Roca de Larios, Murphy, & Manchon, 1999; Wong, 2005), lexical acquisition (e.g., De Bot, Paribakht, & Wesche, 1997; Morrison, 1996; Nassaji, 2003), L1-L2 interface, including translation (e.g., Fraser, 1993; Jaaskelainen, 2000; Seguinot, 1996), computer-mediated communication (e.g., Sachs & Suh, 2007; Smith, 2003), and attention and awareness (e.g., Alanen, 1995; Leow, 1997, 1998a, 1998b, 2000, 2001a, 2001b; Martínez-Fernández, 2008; Rosa & Leow, 2004a, 2004b; Rosa & O’Neill, 1999; Sachs & Suh, 2007).

Although the think-aloud protocols have merits and that they have been widely used in L2 research, their use have been criticized in terms of validity issues (e.g., Russo, Johnson, & Stephens, 1989). The two main concerns that have been put forward with regard to the use of concurrent verbal protocols are non-veridicality and reactivity (e.g., R. Ellis, 2001; Jourdenais, 2001; Nisbett & Wilson, 1977; Payne, Braunstein, & Carroll, 1978). Veridicality concerns the preciseness of verbalization as a representation of participants’ underlying cognitive processes; concurrent verbal reports are non-veridical if they fail to accurately reflect the primary thought process. Reactivity concerns the effect of verbal protocols on participants’ primary cognitive processes, task performance, and possible learning outcomes; concurrent verbal reports are reactive if they alter cognitive processing during on-line task
performance and consequently the end product of that processing.

To empirically test the veridicality of think-aloud protocols poses a methodological challenge because it requires a measure that provides a complete record of information regarding the on-line cognitive processes. Think-aloud protocols are the closest to such a method at present, however. Nonetheless, given that the issue of reactivity precedes the issue of veridicality logically, there is no point of investigating veridicality if the think-aloud procedure has already changed the primary thought processes being reported (Russo et al., 1989, p. 760).

In the field of SLA, it has been cautioned that having learners produce a concurrent verbal report while interacting with L2 input places an additional cognitive processing load on learners and creates a sort of dual task condition, thus not providing a straight measure of their thought processes (Ellis, 2001; Jourdenais, 2001). For instance, Jourdenais (2001) states that “the think aloud data collection method itself acts as an additional task which must be considered carefully when examining learner performance” (p. 373). However, these arguments are not based on empirical findings in L2 research. The next section reviews empirical studies addressing the reactivity of think-aloud protocols in SLA.

2.2.3.2. Empirical Studies on Reactivity in SLA

While empirical investigations into the potential reactivity of think-aloud protocols have been conducted since the 1950s in cognitive psychology, reactivity has only very recently been explored within the field of SLA. To date, six published
studies have empirically addressed the methodological issue of reactivity in L2 research (Bowles, 2008; Bowles & Leow, 2005; Leow & Morgan-Short, 2004; Sachs & Polio, 2007; Sachs & Suh, 2007; Sanz, Lin, Lado, Bowden, & Stafford, 2008).

Leow and Morgan-Short (2004) were the first to carry out empirical research investigating the effects of think-aloud protocols on L2 task performance. In this study, 77 first-year college-level learners of L2 Spanish were randomly assigned to two groups: think-aloud and non-think-aloud. Participants in the think-aloud condition were asked to think aloud as naturally as they could while reading a Spanish text whose targeted linguistic form was a morphological structure, the formal imperative in Spanish, and completing post-exposure assessment tasks. On the other hand, the non-think-aloud group participants completed the reading task and subsequent post-exposure tasks silently. Non-reactivity was found for concurrent verbal reports, because thinking aloud during an L2 reading task did not have any significant, either positive or negative, effects on learners’ text comprehension, recognition or controlled written production when compared to the non-think-aloud group performing the same tasks.

Expanding the scope of work by Leow & Morgan-Short (2004), Bowles and Leow (2005) explored the reactivity of two types of concurrent verbal protocols: non-metalinguistic verbalization and metalinguistic verbalization. While using the same type of an L2 reading task, this study used a syntactic structure and advanced learners of L2 Spanish. Forty-five fifth-semester Spanish learners read a reading passage targeting the
Spanish pluperfect subjunctive in one of three conditions: metalinguistic think-aloud, non-metalinguistic think-aloud, and control (silent). Participants in the metalinguistic group were instructed to “say whatever passes through your mind,” while those in the non-metalinguistic group were directed to verbalize not only their thoughts per se but also justifications and explanations of their reasoning. Results suggested that, compared to a control group, neither type of verbalization significantly impacted on reading comprehension or written production of old or new exemplars of the targeted structure. However, the metalinguistic group performed significantly worse on reading comprehension task than the non-metalinguistic group. In addition, in terms of the amount of time spent in processing the L2 reading passage, it was found that both verbalization groups took significantly longer than the control group.

Bowles (2008) sought to further elucidate the methodological issue of reactivity in SLA by investigating the effects of type of verbalization on controlled written production of previously encountered and novel items of a targeted structure on an L2 problem-solving task. The study also addressed the possible interaction between the type of feedback (implicit vs. explicit) and the verbalization condition. One hundred ninety-four beginning learners of L2 Spanish completed a series of computerized mazes in one of the six following conditions: metalinguistic-implicit, metalinguistic-explicit, non-metalinguistic-implicit, non-metalinguistic-explicit, silent-implicit, and silent-explicit. Contrary to Bowles & Leow (2005), it was reported that whereas metalinguistic verbalization significantly lengthens the time
learners spent performing the task and appears to have a negative effect on
production of previously encountered items (i.e., item learning), simply thinking
aloud was not found to be reactive for both item and system learning and for time on
task. Non-reactivity was found for metalinguistic verbalization with regard to the
production of novel items (i.e., system learning), and there was no interaction
between verbalization and feedback.

Sachs and Polio (2007) found contradicting results from those of Leow &
Morgan-Short (2004), Bowles & Leow (2005), and Bowles (2008). This study set
out to mainly investigate the relative effects of error corrections versus
reformulations on ESL learners’ linguistic accuracy on a subsequent revision of their
composition. Sachs and Polio incorporated think-alouds in two experiments reported
in this paper to examine learners’ reported awareness while comparing their original
stories with the more targetlike reformulations. For the purpose of addressing the
potential effects of thinking aloud as a research instrument, one of the two groups in
the reformation condition were asked to think aloud in English, the L2, while the
other group processed the reformulations they received silently. Reactivity was
found for linguistic accuracy in the first study, but not in the second. In other words,
participants who thought aloud while comparing their composition to the
reformulated version yielded significantly less accurate revisions than those who
processed the feedback silently in one of the two experiments. However, the
differing results from the two studies were not addressed. In their second study
which found no reactivity, different levels of L2 proficiency of the participants who verbalized may have affected the results given that they thought aloud in their L2. In other words, those with a higher level of L2 ability might have evened out any possible effect of think-aloud protocols.

Sachs and Suh (2007) incorporated think-aloud protocols in their investigation into the effects of textually enhanced recasts on L2 development in relation to learners’ reported awareness in synchronous computer-mediated interaction. They addressed the methodological concern that think-alouds may be reactive as a research tool, although this issue was not included in the research questions. Thirty Korean EFL learners of intermediate to high-intermediate level engaged in one-on-one written computer-mediated communication with a native English-speaker interlocutor. They recounted a story they had just read in their L1 in one of four conditions, depending on whether they received textually enhanced or unenhanced recasts and whether they were asked to think aloud during the interactive treatment. Immediately after the treatment, learners completed two types of posttests: a multiple-choice text completion, designed to assess learners’ ability to select the targeted structure, the backshifting of verbs, and an interactive computer-mediated test, designed to measure learners’ ability to produce the target in written communication. Carrying out a repeated-measures ANOVAs, the researchers found a significant between-group main effect for think-alouds on text completion, indicating that the think-aloud and non-think-aloud groups performed differently
from each other regardless of time. However, there was no significant interaction between think-aloud condition and time (pre versus post), which led Sachs and Suh to conclude that thinking aloud did not affect the learners’ improvement from pre- to posttest. Although strong implications regarding reactivity were not made based on the statistical tests, some of the participants reported in a post-study debriefing questionnaire that verbalizing their thoughts had been difficult or distracting.

In a recent publication, Sanz et al. (2008) reported two experimental studies on reactivity, which differed only in the explicitness of the input (i.e., the provision of metalinguistic grammar lesson) and the third choice provided in the assessment materials. Sanz et al. investigated the effects of concurrent verbalizations on the assignment of semantic function to noun phrases in Latin and on latency which they operationalized as the mean reaction time of correct responses on each assessment test. In the first experiment, 24 native English-speaking college-age students with no prior knowledge of a case-marking language engaged in a series of computerized lessons either in think-aloud or silent condition. The treatment consisted of vocabulary presentation, individual and interactive grammar lesson with practice and feedback, and task-essential practice with explicit feedback. While performing the grammar lesson and practice with explicit feedback, think-aloud group learners were asked to say out loud in their L1, English, everything that they would say to themselves silently while they think, but not trying to explain their thoughts. Immediately following the treatment, participants completed three posttests
consisting of an aural interpretation test, a written grammaticality judgment test, and a sentence production test. Separate repeated measures ANOVAs run on the accuracy scores from each of the three tests revealed no significant difference between the think-aloud and silent groups. However, in terms of latency, concurrent verbalization significantly increased reaction time on the grammaticality judgment posttest.

In the second experiment, another group of 24 learners who met the same selection criteria as in the first experiment completed the treatment in the think-aloud or silent condition. Learners in both groups engaged in a vocabulary learning session, and task-essential practice with explicit feedback. Since there was no explicit grammar lesson in this experiment, to control for amount of information provided and time-on-task, Sanz et al. included two additional sets of practice with explicit feedback. Repeated measures ANOVAs revealed significant interactions between think-aloud condition and time (pre versus post) on two of the three tests: the grammaticality judgment test and the production test, which suggests that the think-aloud group improved significantly more than the silent group. Regarding reaction times, no reactivity was found on any of the assessment tests. Combining the results of the two studies, the researchers concluded that concurrent verbalizations have the potential to change the cognitive processes they are meant to reflect, and furthermore, that thinking aloud may induce reactivity depending on the type of task (explicitness in the input) and on the nature of the assessment tool.
In sum, five studies found no reactivity between silent control and verbalization groups (Bowles, 2008; Bowles & Leow, 2005, Leow & Morgan-Short, 2004; Sachs & Suh, 2007; Sanz et al., 2008 (Exp.1)). However, a difference was found between metalinguistic and non-metalinguistic verbalization groups (Bowles & Leow, 2005), and two found think-alouds to be reactive when compared to a silent group (Sachs & Polio, 2007; Sanz et al., 2008 (Exp.2)). A difference was also shown between metalinguistic verbalization and silent groups on item-learning (Bowles, 2008).

The results from previous research together appear to indicate that the effect of think-aloud protocols may depend on the type of task (e.g., reading, writing, problem-solving task, more-less explicit), the type of verbalization prompted (i.e., metalinguistic or non-metalinguistic), the language of verbalization (i.e., L1 or L2), the measure of reactivity (e.g., accuracy, time on task, reaction time), the measurement task of L2 learning (e.g., recognition test, production test, interpretation test, grammaticality judgment test), and the difficulty of the linguistic targets.

Nonetheless, some points need to be taken into account when comparing these studies and synthesizing the findings. In Leow and Morgan-Short (2004) and Bowles and Leow (2005), participants were asked to verbalize their thoughts during the treatment and the immediate posttests, whereas concurrent verbal protocols were yielded only during the treatment in the other SLA studies. In order to determine the
effect of think-aloud protocols in comparison with silent control on the assessment
tests (i.e., measures of L2 learning), verbalization data should be sought during
treatment only. As Leow and Morgan-Short acknowledge (p. 51), their findings of
no reactivity do not suggest that concurrent verbalization does not influence
participants’ performance on assessment tests under comparable conditions.

Next, when addressing reactivity in terms of accuracy, some of these studies
have used the data from immediate posttests (Bowles, 2008; Bowles & Leow, 2005;
Leow & Morgan-Short, 2004), while others have examined gain scores from pretest
to posttest using repeated-measures ANOVAs (Sachs & Polio, 2007; Sachs & Suh,
2007, Sanz et al., 2008). One of the main purposes addressing the issue of reactivity
in SLA is to unveil whether think-aloud protocols affect learners’ internal cognitive
processes to a certain extent to that they alter L2 learning outcomes when compared
to silent condition. In light of this, it seems important to examine the possible
reactivity of concurrent verbalizations with respect to learners’ change from pretests
to posttests.

In sum, no definitive conclusions can as yet be drawn based upon this small,
mixed body of research, given the many potential variables that may affect the
reactivity and the inconclusive findings in cognitive psychology, where research on
reactivity has been carried out for over 50 years. More research on the reactivity
issue of think-aloud protocols in the field of SLA is clearly warranted. Therefore,
while this dissertation incorporates think-aloud protocols as a data collection method
as an attempt to directly measure the constructs of attention and awareness, the reactivity issue is also addressed.

2.3. Rationale and Research Questions

The review of literature presented in the previous sections indicates some areas in need of further research in the field of SLA. As the literature on the role of written feedback in L2 development suggests, the effectiveness of written feedback has received relatively limited attention from researchers. Moreover, studies have not always provided clear-cut and consistent results. Only a few studies (e.g., Bitchener, 2008, Bitchener & Knoch, 2008; Sheen, 2007) have yielded positive evidence for the facilitative effects of written feedback on L2 learning. Furthermore, several written feedback studies have sought to determine the relative effectiveness of different types of written feedback, in particular, direct and indirect feedback, with mixed results. Therefore, it is clear that further research is warranted in order to investigate the link between written feedback and L2 learning and to ascertain which feedback type is most effective.

In addition to the research design and methodological limitations of previous empirical research, it has been suggested in several strands of SLA research, that a possible explanation for the inconsistent results of previous research may pertain to the type of linguistic targets (e.g., Bitchener et al., 2005; Leow et al., 2003; Long et al., 1998; Mackey, 2006; Mackey et al., 2000; Nagata & Swisher, 1995; Shook, 1994; VanPatten, 1990). To address this issue, the present study attempts to
investigate the possible role of type of linguistic target by choosing two linguistic structures that differ in their linguistic salience.

A major underlying assumption of written feedback research is that written feedback provided on learners’ written production would draw learners’ attention to the features marked/corrected, thereby promoting superior, subsequent awareness and further processing of the attended features, and would impact learning when compared to a no feedback condition. Given these premises concerning attention and awareness, it is important to operationalize and measure them in as accurate a manner as possible. However, prior studies on the developmental benefits of written feedback have only made an indirect inference as to the presence and the focus of attention and awareness together with learners’ subsequent processing of L2 data by means of post-exposure assessments. To fill this gap, this study uses introspective methods of exploring attentional processes (e.g., think-aloud protocols) to collect more direct evidence of how the cognitive processes learners are engaged in during exposure to different types of written feedback regarding different types of linguistic targets work and how those processes relate to L2 development.

While employing think-aloud protocols as a data collection method in an effort to directly measure the constructs of attention and awareness, this study also explores the methodological issue of the reactivity, keeping in mind that concerns have been raised as to the potentially reactive nature of concurrent verbal reports (e.g., Jourdenais, 2001) and more empirical research investigating this issue is
needed in SLA.

Specifically, this study was guided by the following research questions:

RQ 1. Does the type of written feedback (indirect vs. direct vs. no feedback) impact learners’ development of linguistic targets? If so, does this effect last over one week?

RQ 2. Does the type of written feedback impact learners’ reported awareness of the linguistic targets at different levels?

RQ 3. Is there a relationship between learners’ reported awareness of the linguistic targets at different levels and their development?

RQ 4. Does verbalization (think-aloud vs. non-think-aloud) impact learners’ development of linguistic targets?
CHAPTER 3: RESEARCH DESIGN AND METHODOLOGY

This chapter reports details of the methodology that was employed to answer the research questions. The chapter first describes the research design and then provides information concerning the study’s participants, the targeted linguistic structures, the operationalizations of direct and indirect written feedback, and the treatment and assessment materials. This is followed by an explanation of the experimental procedure, scoring procedures, and the statistical analyses employed.

3.1. Research Design

This study employed a pretest-treatment-immediate posttest-delayed posttest design to investigate the effects of (differential) written feedback on L2 learners’ development of two linguistic targets: English past counterfactual conditional and the objective-of-preposition (OPREP) type of relative clauses. Participants were randomly assigned to one of the five experimental conditions which differed with respect to the three variables: (1) whether the learners received written feedback, (2) what type of written feedback they received (direct vs. indirect), and (3) whether or not they were asked to think aloud during the feedback session (+/- TA). Participants in the control group did not receive any feedback and took part in this experimental study silently. In other words, the five groups that participated in this experimental study were as follows: (1) Direct Feedback + Think-alouds group, (2) Direct feedback – Think-alouds group, (3) Indirect Feedback + Think-alouds group, (4) Indirect feedback – Think-alouds group, and (5) Control group. The main
The independent variable under investigation was written feedback condition.

The pre- and post-tests were designed to assess the participants’ ability to recognize the appropriate use of the target structures and to produce them accurately. Think-aloud protocols and post-exposure debriefing questionnaires were used to examine whether type of feedback was related to the participants’ reported awareness of the linguistic targets during the feedback session. Several inferential statistical methods (e.g., repeated-measures ANOVAs, chi-squares) were utilized to investigate relationships between the participants’ reported awareness at different levels and their gain scores from the pre- to post-tests. Two linguistic structures were targeted (1) to investigate whether the effect of written feedback varies depending on the type of linguistic target, and (2) to explore relationships among type of written feedback, the participants’ reported awareness, and learning outcomes in relation to the type of linguistic targets.

In order to investigate whether there was any difference in participants’ improvement of accuracy in recognizing and producing linguistic targets when they thought aloud during a feedback session versus when they did not, inferential statistics were employed.

3.2. Participants

The participants in this study were 81 out of an original total of 102 Korean adult EFL learners attending a university in Seoul, South Korea. Most of the participants were enrolled in first-semester college English classes at the university.
during Spring or Summer 2008. To maximize the homogeneity of the sample, only 81 participants fulfilled the following criteria to be selected for the final analysis: (a) they had to be native speakers of Korean, (b) they had to demonstrate zero to limited ability to recognize and produce the target structures, (c) they had not spent more than one year living in an English speaking country, (d) they had not received any additional information on the target structures or forms outside the experiment, as stated in the debriefing questionnaire, and (e) they had to complete all sessions of the study.

The first criterion was established in order to ensure that knowledge from the learners’ first language would not provide direct equivalents to the target structures and consequently affect the results of the experiment. It was also considered that different characteristics of different first languages might play a role in what learners notice in the L2 and how they processed and learned the target structures, which, in turn, would influence the results of the study. Participants’ first language was determined with the aid of a background questionnaire (see Appendix A).

The rationale for the second criterion was the fact that it is essential that participants demonstrate no prior mastery of the target structures in order to investigate language development subsequent to a certain treatment. This was assessed by means of two pre-tests, namely a picture prompt written production test and a recognition/interpretation test. As an effort to prevent the target structures from being obvious to participants on the pre-tests, a series of distractors were
included in the tests. In order to demonstrate zero to limited familiarity with using the target linguistic structures, participants had to score no higher than 50% on the written production of the past counterfactual conditional \((n = 65)\). A score of below 67% on the production pretest was required for the OPREP types of relative clauses \((n = 58)\). The cut-off score was initially set at 50% for the production of OPREP types of relative clauses as well, but was raised to 67% to preserve the sample size. In terms of the recognition test, it was deemed necessary to raise the original, moderately low cut-off score to 80% to secure a sound number in each cell that would allow for inferential statistical analyses \((n = 61\) for past counterfactual conditional, \(n = 41\) for OPREP relatives).

The third criterion was included to make sure that context of second language learning (second language context vs. foreign language context) would not play a role in learning of target structures. This was determined by means of the background questionnaire.

The fourth criterion was established to ensure that post-treatment differences between groups were due to respective treatments that each group received and not to external variables such as outside exposure (e.g., a grammar book, a friend) to the target structures during the experiment. Data for this were gleaned from a post-treatment debriefing questionnaire (see Appendix B).

The final pool consisted of 36 female and 45 male participants. Their ages ranged from 19-30 years, with an average of 24 years. Most of them had begun
studying English in middle school (at around age 13) or in 3\textsuperscript{rd} grade of elementary school (at around age 9), and they had an average of 10 years of previous English study (range 3-20 years). However, the participants reported that they used English for an average of 5 hours per week (range 0.5-40 hours). The majority of the learners had never lived in a country where English was spoken as a native language. Only 7 participants had experienced living in an English-speaking country, and their length of residence in an English-speaking country was less than 11 months. About half of the participants had experience with other foreign languages, with 17 reporting at least basic abilities in Japanese, 16 in Chinese, 10 in Spanish, and 5 in German. Forty two participants had no experience with other languages.

Participants were randomly assigned to one of four experimental conditions and a control condition created for this experiment, and given an informed consent form to complete. The researcher made every effort to protect participants. All participants remained anonymous to the fullest extent possible. Once the data were collected, their names were not used and ID codes were given.

3.3. The Targeted Linguistic Structures

The two linguistic structures targeted in the present study were the past counterfactual conditional in English (e.g., \textit{If I had not overslept, I would not have missed the bus yesterday.}), and the objective-of-preposition (OPREP) types of relative clauses in English (e.g., \textit{The man who(m) I am looking for is tall.}).

In English, the imaginative (or subjunctive) conditional sentences can be
categorized into two subtypes: hypotheticals (e.g., If Susan had the time, she would go to Washington D.C.) and counterfactuals. Counterfactual conditionals are further subcategorized into the present counterfactuals (e.g., If my grandmother were alive today, she would see a different world.) and the past counterfactuals (Celce-Murcia & Larsen-Freeman, 1999, p. 551). The past counterfactual conditional describes impossibilities in reference to the past (e.g., If my grandmother were alive in 2005, she would have been 88).

As noted in Celce-Murcia & Larsen-Freeman (1999), the (past) counterfactual conditional is considered as one of the most problematic features for L2 learners of English, mainly due to its (i) syntactic complexity, consisting of two clauses, a subordinate clause and a main clause, and taking certain modal verbs in main clauses, and (ii) semantic difficulty, subtle meaning distinction among various types of conditional clauses. In the imaginative conditionals, the past tense refers to the present time, and the past perfect tense refers to past time, which often causes difficulties even for high-intermediate to advanced L2 learners.

In addition, some differences between the past counterfactual conditional in English and its counterpart in Korean may cause substantial difficulty for Korean learners of English (Cowan, 2008). A major difference in conveying conditional meanings between the two languages appears to relate to the dissimilar number of tenses used in conditional sentences. In Korean, there are a relatively limited number of tenses employed in conditionals. In English, on the other hand, a larger number of
tense sequences are used to describe specific conditional meanings. Another L1-L2 difference pertains to the fact that, unlike in English, past tense morphemes occur in both clauses of counterfactual conditionals in Korean. Korean learners of English, therefore, seem to have difficulty acquiring appropriate tense sequences required for English conditionals. In fact, Korean learners often produce “what appear to be hypothetical conditionals in which the context of the discourse clearly indicates the need for a counterfactual conditional” as in “If I was not forced to learn, I would never have chance to learn English” (Cowan, 2008, p.426).

Another linguistic target, the OPREP types of relatives, was chosen in consideration of Izumi’s (2000, reported on 2002) research on English relativization which draws upon Keenan and Comrie’s (1977) Noun Phrase Accessibility Hierarchy (NPAH) and Doughty (1988, 1991). According to NPAH, it is suggested that the English OPREP relatives are more difficult to process than relative clauses embedded in subjective (SU), direct object (DO), and indirect object positions (IO). The pre-test results in this study seem to concur with this claim, given that participants often showed a correct recognition and/or use of SU or DO types of relatives, even when they revealed problems with recognition and/or production of OPREP types of relative clauses. For L2 learners of English, the relative clauses may pose difficulty because of its complexity which involves in (i) the position of the relatives with respect to the head noun (i.e., relatives occur before or after the head noun), (ii) the way of marking relative clauses (e.g., who in English), and (iii)
the presence and absence of a pronominal reflex (e.g., *The man who I am looking for him is tall) (Celce-Murcia & Larsen-Freeman, 1999, p. 575).

Korean learners of English are expected to experience difficulty in learning the relative clauses, given several differences in the grammaticalization of relatives in English and Korean. The important characteristics of relative clauses in Korean encompass (i) its head-finality (i.e., relatives precedes the head noun) and basic constituent order, subject-object-verb (Lee, 2007), (ii) the lack of relative pronoun, and (iii) the use of “a set of adnominal verbal suffixes that express the tense of the embedded clause: -(u)n, -nun, and –(u)l” (Jeon & Kim, 2007, p. 255).

In addition to their linguistic difficulty for (high) intermediate and even advanced Korean EFL learners, the rationale for selecting these structures as the linguistic targets relates to the assumption that they seem to vary in linguistic salience. Although both structures have marks (e.g., ‘if,’ ‘who’), relative clauses are usually broader in scope than the conditionals. It may be, therefore, that the past perfect form and modal verbs make the past counterfactual conditional a perceptually more salient structure to learners. It was also reasoned that the two linguistic structures would possibly be learnable through three treatment sessions with focused feedback, given both the proficiency level of learners in this study and the promising results from previous research on learning of these forms in the instructional setting (e.g., Izumi & Bigelow, 1999; Izumi, 2002; Song & Suh, 2008).
3.4. Operationalizations of Direct and Indirect Written Feedback

In order to examine whether different types of feedback have different learning effects, depending on the directness or explicitness with which learners’ errors are corrected, two forms of written feedback were manipulated in this research. Direct feedback was operationalized as indication of linguistic error through underlining and/or cursors (∧) and provision of correct linguistic forms or structures above the learners’ incorrect use of the targeted structures. Indirect feedback indicates the existence and location of learner errors on the learner’s writing by means of underlining and/or cursors (∧). Therefore, the major difference between the two types of written feedback lay in the presence and absence of correct form/structure, or positive evidence. It was postulated that learners in the indirect feedback condition would engage in problem solving to correct an error, when provided indirect feedback.

Below, (1) and (2) are typical examples of direct feedback taken from the data of the present study, while (3) and (4) show typical examples of indirect feedback. The actual feedback was provided using a red pen.

**Examples 1-2. Direct feedback**

would not have missed
(1) If Lisa had not overslept, she didn’t miss the bus.
person who she was fond of

(2) She wanted to have fun with the person who was fond of, so ……

**Examples 3-4. Indirect feedback**

(3) If Lisa had not overslept, she didn’t miss the bus.

(4) She wanted to have fun with the person who was fond of, so ……

As can be seen in (2), other linguistic errors such as erroneous article use were not corrected for the purposes of the present study. That is, the definite article “the” that precedes the head noun of the OPREP types of relative clauses was not underlined, nor was corrected as “a,” a more appropriate form given the context. The control group in this study received no feedback.

3.5. Materials

3.5.1. Treatment Materials

The treatment consisted of three experimental treatment sessions, each of which consisted of a 2-day sequence of composition and feedback. A written story-retelling task was used for the treatment. Six versions of the story-retellings were created to (1) elicit production of the two linguistic targets: past counterfactual conditional and OPREP types of relative clauses in English, and (2) provide opportunities for the written feedback described above to take place. Three of them served as treatment tasks, and the other three served as production assessment tests.
All six stories included five main clauses and five if-clauses, which compose five target counterfactual conditional sentences, and five OPREP relatives. The stories were adapted from Izumi (2000), Song & Suh (2008), and Suh (2007). The story-retelling tasks were guided by picture prompts. The participants were given fourteen through sixteen sequenced picture and vocabulary prompts which were designed to elicit the targets together with four through six distractors. The rationale for providing picture and vocabulary prompts was to (1) lessen the memory load of remembering a story, and (2) keep the content and length of learners’ written production as well as the amount of use of linguistic targets similar and comparable across the groups. A small-scale pilot study which was conducted on a separate pool of learners similar to those in the main study showed that six versions of the story-retelling task were comparable.

In the instructions given prior to the task, participants were told that they were going to read a story in Korean and retell the story in English in the written form on a blank sheet of paper. They were also informed that they would be provided a series of sequenced cards displaying pictures and vocabulary words. The learners were asked to write one sentence per each picture card to complete an entire story, and they were encouraged to use the words written below each picture to write each sentence. The purpose was to make sure that all the target sentences would be attempted and that the participants would be given contextualized feedback on the same structures. All group learners had 30 minutes to complete the writing.
During the treatment, learners read a short story in Korean silently and placed it aside. All group participants were then given a set of sequenced cards depicting the main events of the story in the form of pictures along with vocabulary items in English so that they could retell the story in English. Appendix C contains the six sets of picture prompts. The participants’ written narratives were collected upon completion. The researcher provided different types of written feedback on errors involved in the two linguistic structures according to their group assignment. Each group learners underwent the feedback stages under each of the five conditions: (1) Direct Feedback + Think-alouds, (2) Direct Feedback – Think-alouds, (3) Indirect Feedback + Think-alouds, (4) Indirect Feedback – Think-alouds, and (5) Control. Therefore, the only difference among the conditions took place on the second day of the sequence, during the feedback stage. The learners were told that they could not take any notes during the feedback stages.

3.5.2. Assessment Materials

Two types of pre- and posttests were developed: (1) target structure recognition/interpretation tests, devised to assess the learners’ ability to recognize the appropriate uses of the target structures, and (2) written production tests, devised to assess the ability to produce the past counterfactual conditional and the OPREP types of relative clauses appropriately in context. In all the assessment test sessions, a production test was administered before two recognition/interpretation tests so that
learners were not provided with additional input from the multiple-choice recognition test items, which could in turn affect their performance in the production test. Both tests were administered in a paper-and-pencil format.

3.5.2.1. Recognition/Interpretation Test

In order to assess the learners’ ability to choose appropriate uses of the English past counterfactual conditional before and after the treatments, multiple-choice (MC) recognition tests were employed. The MC recognition tests contained twenty blanks, of which ten were targets and ten were distractors requiring articles, phrasal verbs, and active and passive voice participles as adjectives. Of the ten requiring the target structure, five were embedded in main clauses, while the other five were embedded in if-clauses. The learners were given four choices for each blank (see Appendix D). Three versions of the MC recognition tests were designed and they were counterbalanced for order across groups.

To assess learners’ ability to recognize/comprehend appropriate uses of the OPREP types of relative clauses, a picture-drawing task test was designed. This test was inspired by assessment tasks using pictures in other SLA studies (e.g., Izumi, 2002; Mackey, 1999; Sanz & Morgan-Short, 2004; VanPatten & Cadierno, 1993), and the test items were adapted from those used in Izumi (2000)’s interpretation test. The test included 10 sentences, of which five were OPREP relatives and the other five were other types of relatives chosen as distractors. The five distractor sentences consisted of three relative clauses embedded in the subject position, and two
relatives embedded in the object position. Participants were told to read each sentence given on their test sheet and draw a simple picture describing the sentence. They were given 10 minutes to complete the test. The same test items were utilized in pretest and posttests, however, the items were presented in a different order from test to test.

3.5.2.2. Production Test

As mentioned in the previous section, three of the six guided story-retellings served as the production tests. The three versions of the story-retelling tasks were counterbalanced to prevent possible order effects. For all the story retelling tasks, picture prompts were provided. The first, fifth, and sixth learner compositions (without providing feedback or verbalization) served as pre-, immediate post- and delayed post-tests of production. Again, all stories contained five target counterfactual conditional sentences, which consist of five main clauses and five if-clauses, as well as five OPREP relatives.

3.5.2.3. Measures of Awareness

It is important to keep in mind, as pointed out by other researchers (e.g., Izumi, 2002; Izumi & Bigelow, 2000; Mackey, 2006; Rosa & Leow, 2004a), the importance of employing multiple measures in exploring complex notions such as attention and awareness. In light of this, the current study implemented a triangulation of measures: online think-aloud protocols concurrent to the feedback session and the retrospective debriefing questionnaire completed after the delayed
posttest. Nevertheless, as discussed in Chapter 3, the online nature of the think-aloud protocols makes it currently the most direct and reliable source of cognitive data, and has proved to be fruitful in investigating learners’ internal processes of L2 input and distinguishing between their levels of awareness (e.g., Leow, 1997, 2000; Rosa & Leow, 2004a; Rosa & O’Neill, 1999; Sachs & Suh, 2007), in line with Schmidt’s (e.g., 1990, 1995) distinction between two levels of awareness. Therefore, in this study, think-aloud protocols were used as the primary source for coding participants’ different levels of reported awareness. Only when verbal report of awareness of the target structures was not available, the answers to the debriefing questionnaire were used in an attempt to classify learners more accurately with respect to the highest levels of awareness they reached during the treatment.

3.5.2.3.1. Think-aloud protocols.

Prior to the feedback session of each treatment, the learners in the think-aloud condition were asked to speak aloud any thoughts that naturally ran through their minds, while reading aloud their writings with feedback provided. Although the learners were told to think aloud either in their L1, Korean, or in L2, English, in whichever language they felt more comfortable, everyone thought aloud in Korean except that when they read aloud their L2 output or feedback. They were told to verbalize their thoughts as naturally as possible, but were not asked to provide metalinguistic information. The instructions were adapted from Sachs and Suh (2007). Before the first feedback session, as a practice, the participants talked aloud
while solving three simple math problems (e.g., $15 + 8 \times 9$). The think-aloud data were recorded for subsequent coding.

3.5.2.3.2. Debriefing questionnaire.

The debriefing questionnaire asked learners whether they had inferred anything about the purpose of the overall experimental tasks, whether they thought the researcher had been focusing on something specific when she had given feedback, if so, what they thought she had been focusing on, whether they had noticed or learned anything interesting about English during feedback session, whether they could state any rules they might have come up with, and how they felt about thinking aloud during the feedback session (Appendix B). Two versions of the debriefing questionnaire were created for the think-aloud group, and the non-think-aloud and control groups. Same questions were included in each version of the questionnaire, but two additional questions were added to the version for the think-aloud group.

3.6. Procedure

An overview of the experimental procedure for this study can be seen in Figure 1. The experiment was carried out over the course of four weeks. On the first day of the experiment, the learners took two types of pretest: (1) a production test (in the form of a story-retelling task), and (2) a recognition test consisting of multiple-choice questions for the past conditional and a picture drawing task for the relative clauses. They spent 30 minutes on the production test and 15 minutes on the
recognition test. Then, the participants completed their second written story-retelling task, and handed in the compositions to the researcher. On the second day of the experiment, which typically occurred 2 to 4 days later, the learners participated in a feedback session during which they received their second stories under one of the following conditions:

1. **Direct feedback + Think-alouds**: Received their stories with errors underlined/marked and with corrected form or structure provided above their errors, and looked at them while thinking aloud.

2. **Direct Feedback – Think-alouds**: Received their stories with errors underlined/marked together with corrections, and looked at them silently.

3. **Indirect Feedback + Think-alouds**: Received their stories with errors underlined/marked, and looked at them while thinking aloud.

4. **Indirect Feedback – Think-alouds**: Received their stories with errors underlined/marked, and looked at them silently.

5. **Control**: Received their stories with no feedback provided, and looked at the original stories with no feedback provided.

They were given approximately 5 minutes to look at their stories (with feedback) during the feedback stage, and they had to return the stories to the researcher. The third written story-retelling task followed. On the third day of the experiment, they took part in the second feedback session during which they received the third stories under each of the five conditions: 1) Direct Feedback + TA,
2) Direct Feedback – TA, 3) Indirect Feedback + TA, 4) Indirect Feedback – TA, 5) Control. After returning their stories, they completed the fourth written story-retelling task. On the Day 4 of the experiment, the participants underwent the third feedback session under one of the five conditions. Then, they took the two types of immediate posttest: a production test and a recognition test. Finally, approximately a week after the third feedback treatment, all the participants took the two types of delayed posttest and filled out a debriefing questionnaire (based on Rosa & Leow, 2004).

The participants’ stories produced during the three treatment sessions were examined in order to verify whether the tasks were successful in eliciting the target structures and whether the experimental conditions were similar except the planned differences in feedback type. As shown in Table 1, participants who received different type of feedback made similar mean number of attempts at the target sentences. In addition, the Direct Feedback and Indirect Feedback groups did not significantly differ in the number of feedback provided throughout the three feedback sessions.
Figure 1. *Procedure*

<table>
<thead>
<tr>
<th>Treatment 1</th>
<th>Treatment 2</th>
<th>Treatment 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Direct Feedback</td>
<td>Direct Feedback</td>
<td>Direct Feedback</td>
</tr>
<tr>
<td>+ TA</td>
<td>- TA</td>
<td>+ TA</td>
</tr>
<tr>
<td>- TA</td>
<td>+ TA</td>
<td>- TA</td>
</tr>
<tr>
<td>Indirect Feedback</td>
<td>Indirect Feedback</td>
<td>Indirect Feedback</td>
</tr>
<tr>
<td>+ TA</td>
<td>- TA</td>
<td>+ TA</td>
</tr>
<tr>
<td>- TA</td>
<td>+ TA</td>
<td>- TA</td>
</tr>
<tr>
<td>Control</td>
<td>Control</td>
<td>Control</td>
</tr>
</tbody>
</table>

Picture prompt written production pretest

Multiple-choice recognition and picture drawing task pretests

**Figure 1. Procedure**

- Picture prompt written production pretest
- Multiple-choice recognition and picture drawing task pretests
- Direct Feedback + TA, Direct Feedback - TA, Indirect Feedback + TA, Indirect Feedback - TA, Control

83
<table>
<thead>
<tr>
<th>Multiple-choice recognition and picture drawing task immediate posttests</th>
</tr>
</thead>
<tbody>
<tr>
<td>↓↓↓ (1 week interval) ↓↓↓</td>
</tr>
<tr>
<td>Picture prompt written production delayed posttest</td>
</tr>
<tr>
<td>Multiple-choice recognition and picture drawing task delayed posttests</td>
</tr>
<tr>
<td>Debriefing Questionnaire</td>
</tr>
</tbody>
</table>
Table 1

*Number of attempts at target sentences and number of feedback received in each feedback condition*

<table>
<thead>
<tr>
<th></th>
<th>Past Counterfactual Conditional</th>
<th>OPREP Relatives</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Direct Feedback</td>
<td>Indirect Feedback</td>
</tr>
<tr>
<td>Number of attempts</td>
<td>13.0</td>
<td>14.0</td>
</tr>
<tr>
<td>Number of feedback received</td>
<td>6.8</td>
<td>10.0</td>
</tr>
</tbody>
</table>

3.7. Scoring Procedures

3.7.1. Recognition/Interpretation Tests

In the multiple-choice recognition test designed to assess learners’ recognition of the past counterfactual conditional, 1 point was assigned to each correct target response and 0 to each incorrect one. Since *if*-clauses and main clauses of the conditional sentence were scored separately, the maximum raw score possible in the test was 10. The raw scores were converted to percentage scores so that they
could be compared to the scores obtained from the picture-drawing task test created for the relative clauses. Each learner was given a percentage score in the pre-test, and two scores in the immediate and delayed posttests.

In the picture-drawing task test, one point was given for each correct response, and zero point was given for incorrect one. A total number possible was five. Each participant was given a total percentage score in the pre-test, and two scores in the immediate and delayed posttests.

3.7.2. Production Tests

The participants’ written production data were scored using a Target-Like Use (TLU) analysis (Pica, 1984). The production tests were scored separately for each of the two linguistic targets. The participants received 1 point for each targetlike production of the targeted structures in obligatory contexts, and 0 point for each non-targetlike attempt. The ratio scores were converted to percentage scores for ease of comparisons with other scores. In scoring the learners’ production of the OPREP relatives, errors such as articles and spelling were ignored, if they did not relate to the formation of the relative clauses. In scoring the production of the past counterfactual conditional, incorrect morphology (e.g., “catched,” instead of “caught”) was ignored.

3.7.3. Coding of Think-aloud Data

The learners’ online verbal reports gathered through think-alouds were initially
coded based on Schmidt’s (1995) ideas about levels of awareness and Leow (1997)’s coding method. Leow (1997) coded learners’ think-aloud protocols according to three criteria for the presence of awareness to distinguish different levels of awareness. Learners’ verbal reports were coded as the first category, [+cognitive change (+CC), -meta-awareness (-MA), -morphological rule (-MR)], when they indicated cognitive registration of L2 forms, in other words, when they showed some behavioral or cognitive change due to the experience during the treatment (e.g., they produced targetlike L2 forms or made verbal or written corrections), but made no report of being aware of the experience or verbalized no underlying rule. Verbal reports were coded as the second category, [+cognitive change (+CC), +meta-awareness (+MA), -morphological rule (-MR)], when learners self-reported subjective experiences, but did not verbalize any underlying rule. In the third category, [+cognitive change (+CC), +meta-awareness (+MA), +morphological rule (+MR)], learners demonstrated a report of subjective experiences and verbalized underlying rules metalinguistically. In line with Schmidt’s (e.g., 1990, 1995) two levels of awareness, Leow (1997) further classified the first category, [+CC, -MA, -MR], as representing awareness at the level of noticing, and the third category, [+CC, +MA, +MR], as representing awareness at the level of understanding.

In the current investigation, each verbalization related to written feedback was coded in order to classify learners with respect to the highest levels of
awareness they achieved while interacting with the feedback provided. Learners’
think-aloud data were coded separately by targeted linguistic structure. After initial
analyses, the think-aloud data revealed the need to include a fourth level which
represents no verbal report of awareness (cf. Rosa & O’Neill, 1999). There were
some cases in which learners failed to provide evidence for registration of target
structures with awareness. These learners were coded as no verbal report
(abbreviated as [NVR]), [-cognitive change (-CC), -meta-awareness (-MA), -
morphological rule (-MR)].

Furthermore, in applying the four criteria to the data from the present study,
it was deemed necessary to make a few adjustments to the criteria for coding
verbalizations at certain levels with respect to the type of feedback provided. In the
case of direct feedback group, when participants read the researcher’s input
presented in the direct feedback out loud, when they paused after verbalizing the
targets, or repeated (part of) the targets, these were considered as evidence of being
aware of the targeted forms or structures ([+CC, -MA, -MR]) – in other words,
awareness at the level of noticing ([N]). For indirect feedback group learners, when
they produced the target forms or structures with a rising intonation (or in an unsure
tone), in response to the researcher’s indirect feedback, once or twice, but then kept
verbalizing their erroneous forms for the remaining part, this was considered as
evidence of awareness at the level of noticing ([N]). When participants in the
indirect feedback group verbalized self-corrections or self-correction accompanied
by a sign of subjective experiences such as “ah!,” these were considered to represent meta-awareness ([MA]), even if the participants did not make an explicit report of subjective experiences such as “so I have to remember that” (Leow, 1997, p.479). If participants used specific metalinguistic terminology to identify the grammatical features of the targets and/or verbalized a (partial) underlying rule, this kind of verbalization was coded as morphological rule formation ([MR]).

Four categories of reported awareness were, therefore, identified: [NVR] or [-cognitive change, -meta-awareness, -morphological rule], [N] or [+cognitive change, -meta-awareness, -morphological rule], [MA] or [+cognitive change, +meta-awareness, -morphological rule], and [MR] or [+cognitive change, +/-meta-awareness, +morphological rule]. However, for the purposes of statistical analyses (see Chapter 4 on the results for detailed explanation), the two levels of awareness, [MA] and [MR] were later combined as one level, a high level of awareness or [H]. [N] was later re-labeled as a low level of awareness or [L]. Table 2 presents examples of each coding level as revealed by the protocols in the present study.
Table 2

*Examples of Think-aloud Data Coding Categories*

<table>
<thead>
<tr>
<th>Feedback type</th>
<th>Target contexts in writing</th>
<th>Corresponding verbalizations</th>
<th>Coding</th>
</tr>
</thead>
<tbody>
<tr>
<td>Direct feedback</td>
<td>would not have needed P42: she <em>did not need</em> to repeat the course,</td>
<td>“she would not have needed to repeat the course,”</td>
<td>[N] (awareness at the level of noticing (\rightarrow) low level of awareness)</td>
</tr>
<tr>
<td>Indirect Feedback</td>
<td>P63: If they <em>planned</em> activities better,</td>
<td>“If they planned activities better, <em>should it be</em> had planned?” (kept verbalizing his own erroneous output after this one; italicized portion spoken in Korean)</td>
<td>[N] (awareness at the level of noticing (\rightarrow) low level of awareness)</td>
</tr>
<tr>
<td>Direct feedback</td>
<td>could have been P38: This situation <em>could be</em> prevented</td>
<td>“This situation could have been prevented. <em>I guess I don’t know the rule about verb tenses.</em>”</td>
<td>[MA] (awareness at the level of reporting (\rightarrow) low level of awareness)</td>
</tr>
<tr>
<td>Indirect Feedback</td>
<td>P64: If she didn’t oversleep, “If Lisa (pause) had…had not…overslept,” (continued to make self-corrections for the remaining part, and upon interacting with the last feedback provided, said) “ah!”</td>
<td>[MA] (awareness at the level of understanding → high level of awareness)</td>
<td></td>
</tr>
<tr>
<td>Direct Feedback</td>
<td>P5: If she was prepared for the test, “oh! when I make hypotheses about something that happened in the past, I have to use the past perfect.”</td>
<td>[MR] (awareness at the level of understanding → high level of awareness)</td>
<td></td>
</tr>
</tbody>
</table>

The learners’ think-aloud data were transcribed. After having a training session on 5% of the data, the researcher and a second coder independently coded 20% of the verbalizations (translated into English when spoken in Korean) which were randomly chosen. Inter-rater agreement was 95% for levels of reported awareness of the past counterfactual conditional, and 100% for levels of reported awareness of the OPREP relatives. The remaining verbal reports were then coded by the researcher.
3.7.4. Complementary Use of Debriefing Questionnaire

As previously mentioned, a debriefing questionnaire was utilized as a complementary introspective method to think-aloud protocols in this study. In the occasions in which learners failed to verbalize targeted L2 forms or structures and were thus coded as [NVR], their answers to the debriefing questionnaire were cross-checked to ensure if they were in fact unaware of the target structures.

3.8. Statistical Procedures

This section presents an overview of the statistical analyses that were performed in order to answer the research questions. For ease of presentation, each research question is repeated below.

RQ 1. Does type of written feedback (indirect vs. direct vs. no feedback) impact learners’ development of linguistic targets? If so, does this effect last over one week?

In order to answer RQ1, four 3 x 3 repeated-measures Analysis of Variance (ANOVA) with a one between-subject and a one within-subject design were conducted. The between-subject factor was written feedback condition (Indirect vs. Direct vs. Control) and the within-subject factor was time (the scores obtained in the pretest versus those obtained in the immediate and delayed posttest). Separate analyses were carried out on learners’ recognition and production scores as the dependent variables, and individual analyses were run separately for each linguistic target.

RQ 2. Does the type of written feedback impact learners’ reported awareness of the
linguistic targets at different levels?

To answer RQ2 as well as RQ3 below, only the think-aloud groups were considered. Participants were first regrouped into one of the four categories, that is, [MR], [MA], [N], or [NVR], on the basis of their respective think-aloud protocols and post-exposure debriefing questionnaires. Frequency counts of [MR], [MA], [N], and [NVR] participants in each of the two feedback conditions (i.e., Indirect, Direct) were submitted to a chi-square analysis. In this way, it was possible to determine whether the type of written feedback had a differential effect on the levels of learner awareness raised or the type of cognitive processing reported by learners in each condition.

RQ 3. Is there a relationship between learners’ reported awareness of the linguistic targets at different levels and their development?

In order to answer RQ 3, as with RQ 2, only the think-aloud groups were considered. Separate 2 x 3 repeated-measures ANOVAs were performed on learners’ production test scores of the two target structures, with reported awareness ([L] vs. [H] for the past counterfactual conditional, [NVR] vs. [N+], or awareness at least at the level of noticing, for the OPREP types of relative clauses) as the between-subjects grouping factor, time (pretest vs. immediate posttest vs. delayed posttest) as the within-subjects factor.

RQ 4. Does verbalization condition (think-aloud vs. non-think-aloud) during feedback session impact learners’ development of linguistic targets?
In order to answer RQ 4, individual 2 x 3 repeated-measures ANOVAs were conducted, with verbalization condition (+think-aloud vs. –think-aloud) as a between-subjects variable, and time (pretest, immediate posttest, delayed posttest) as a within-subjects variable.

All analyses were performed using the Statistical Package for the Social Science (SPSS). The alpha level was set at 0.05.
CHAPTER 4: RESULTS

This chapter presents the results of the statistical analyses carried out in order to test each research question. First, the results of descriptive and inferential statistics for research question 1, namely, does type of written feedback (indirect vs. direct vs. no feedback) impact learners’ development of linguistic targets? if so, does this effect last over one week?, are reported. Next, the descriptive and statistical data for research question 2, namely, does the type of written feedback impact learners’ reported awareness of the linguistic targets at different levels?, are presented. Then, the results of descriptive and inferential statistics for research question 3, namely, is there a relationship between learners’ reported awareness of the linguistic targets at different levels and their development?, are described. This chapter also provides the results for research question 4, that is, does verbalization (think-aloud vs. non-think-aloud) impact learners’ development of linguistic targets?

4.1. Learning: Results for the Effects of Written Feedback Condition

The first research question identified for this study addressed the differential effect of direct and indirect written feedback on learners’ ability to recognize, comprehend, and produce the target structures, both immediately after treatment and one week later. To answer the research question one (Does the type of written feedback impact learners’ development of linguistic targets? If so, does this effect last over one week?), separate 3 x 3 repeated measures ANOVAs were performed on learners’ scores in the recognition/interpretation and production of English past
counterfactual conditionals and OPREP types of relative clauses, with feedback condition (DF vs. IF vs. Control) as the between-group factor, and time (pre vs. post vs. delayed-post test) as the within-subjects factor. In addition, 2 x 3 repeated measures ANOVAs and means comparisons were conducted as follow-up analyses when a significant interaction effect was found.

4.1.1. Learning of Past Counterfactual Conditionals

4.1.1.1. Results on Recognition Task

Means and standard deviations were computed for feedback condition across time on learners’ scores in recognition of English past counterfactual conditionals. These data are shown in Table 3.

Table 3

Recognition of Past Conditionals: Mean Test Scores by Feedback Condition across Time

<table>
<thead>
<tr>
<th>Condition</th>
<th>N</th>
<th>Pretest Mean</th>
<th>Pretest SD</th>
<th>Immediate Posttest Mean</th>
<th>Immediate Posttest SD</th>
<th>Delayed posttest Mean</th>
<th>Delayed posttest SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>9</td>
<td>55.56</td>
<td>27.89</td>
<td>65.5</td>
<td>38.77</td>
<td>67.78</td>
<td>34.20</td>
</tr>
<tr>
<td>Indirect</td>
<td>28</td>
<td>48.93</td>
<td>29.86</td>
<td>68.57</td>
<td>34.50</td>
<td>79.29</td>
<td>30.30</td>
</tr>
<tr>
<td>Direct</td>
<td>24</td>
<td>50.83</td>
<td>23.58</td>
<td>87.08</td>
<td>18.29</td>
<td>94.58</td>
<td>13.82</td>
</tr>
</tbody>
</table>
Scores (in percentages) on recognition of the past counterfactual conditionals were submitted to a 3 x 3 repeated-measures ANOVA using a one between-subject, one within-subject design. The between-subject factor was written feedback condition (DF, IF, Control), and the within-subject factor was time (pretest, posttest, delayed posttest). Results of the repeated-measures ANOVA revealed a significant main effect for time, $F(2, 58) = 30.59, p = 0.00$, a significant interaction between feedback condition and time, $F(4, 58) = 3.12, p = 0.02$, but no significant main effect for feedback condition, $F(2, 58) = 2.20, p = 0.12$. The results, then, indicated that, regardless of the feedback group, participants significantly improved over time and that the different groups improved to a different extent across time. Figure 2 visually presents this interaction.
In order to determine the magnitude, or strength, of the interaction effect between time and feedback condition and the main effect of time and in the population, it was needed to assess their effect sizes. There are many strength of association measures deemed acceptable for establishing a sample’s effect size such
as Cohen’s $d$, eta-squared, and partial eta-squared. Among these measures, the present study employed partial eta-squared, since the measure is reported in SPSS. For partial eta-squared, .01 is considered a small effect size, .06 a medium effect size, and .14 a large effect size. With a partial eta-squared of 0.10, the significant interaction effect on the recognition test proves to have a medium effect size in the greater population. These results, along with others, are displayed in Table 4.

### Table 4

**Recognition of Past Conditionals: 3 x 3 ANOVA for Feedback Condition by Time**

<table>
<thead>
<tr>
<th>Source</th>
<th>Sums of Squares</th>
<th>Df</th>
<th>Mean</th>
<th>$F$</th>
<th>$p$</th>
<th>Partial $\eta^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Feedback</td>
<td>7014.61</td>
<td>2</td>
<td>3507.31</td>
<td>2.20</td>
<td>0.120</td>
<td>0.07</td>
</tr>
<tr>
<td>Time</td>
<td>21596.45</td>
<td>2</td>
<td>10798.23</td>
<td>30.59</td>
<td>0.000*</td>
<td>0.35</td>
</tr>
<tr>
<td>Time x Feedback</td>
<td>4409.04</td>
<td>4</td>
<td>1102.26</td>
<td>3.12</td>
<td>0.018*</td>
<td>0.10</td>
</tr>
</tbody>
</table>

* $p < .05$

Given the significant interaction effect, three separate 2 x 3 repeated-measures ANOVAs with one between-subject (feedback condition) and one within-subject (time) design were performed to find out exactly which groups’ gain scores (direct vs. control, indirect vs. control, direct vs. indirect) in recognition from pretest.
to delayed posttest significantly differed from each other. The results showed a significant interaction effect only when direct and control conditions were compared: $F(2, 31) = 7.10, p = 0.002$, partial eta-squared = 0.19 (a large effect size). When the indirect and the control groups’ gain scores over time were examined, a post-hoc power analysis revealed that the observed power for the feedback condition effect was 0.28 (much below the recommended level of 0.8), indicating that the lack of statistically significant results may have been attributed to small sample size. These results are reported in Table 5.
Table 5

*Recognition of Past Conditionals: 2 x 3 ANOVA for Feedback Condition by Time*

<table>
<thead>
<tr>
<th>Source</th>
<th>Sums of Squares</th>
<th>df</th>
<th>Mean Squares</th>
<th>F</th>
<th>p</th>
<th>Partial eta²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Direct vs. Control</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Feedback</td>
<td>1383.22</td>
<td>1</td>
<td>1383.22</td>
<td>3.56</td>
<td>0.069</td>
<td>0.10</td>
</tr>
<tr>
<td>Time</td>
<td>11708.67</td>
<td>2</td>
<td>5854.34</td>
<td>22.27</td>
<td>0.000*</td>
<td>0.42</td>
</tr>
<tr>
<td>Time x Feedback</td>
<td>3732.91</td>
<td>2</td>
<td>1866.46</td>
<td>7.10</td>
<td>0.002*</td>
<td>0.19</td>
</tr>
<tr>
<td>Indirect vs. Control</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Feedback</td>
<td>47.19</td>
<td>1</td>
<td>47.19</td>
<td>0.06</td>
<td>0.81</td>
<td>0.00</td>
</tr>
<tr>
<td>Time</td>
<td>6490.83</td>
<td>2</td>
<td>3245.41</td>
<td>7.88</td>
<td>0.001*</td>
<td>0.18</td>
</tr>
<tr>
<td>Time x Feedback</td>
<td>1121.46</td>
<td>2</td>
<td>560.73</td>
<td>1.36</td>
<td>0.26</td>
<td>0.04</td>
</tr>
<tr>
<td>Direct vs. Indirect</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Feedback</td>
<td>1831.50</td>
<td>1</td>
<td>1831.50</td>
<td>3.98</td>
<td>0.052*</td>
<td>0.07</td>
</tr>
<tr>
<td>Time</td>
<td>38543.68</td>
<td>2</td>
<td>19271.84</td>
<td>52.39</td>
<td>0.00*</td>
<td>0.51</td>
</tr>
<tr>
<td>Time x Feedback</td>
<td>2005.22</td>
<td>2</td>
<td>1002.61</td>
<td>2.73</td>
<td>0.07</td>
<td>0.05</td>
</tr>
</tbody>
</table>

* p < .05

Post-hoc means comparisons revealed that the direct feedback group made a
significantly larger improvement than the control group from pre-test to immediate post-test, \( F(1, 31) = 6.31, p = 0.017 \), partial eta-squared = 0.17 (a large effect size), but not from immediate to delayed post-tests, \( F(1, 31) = 0.71, p = 0.407 \), partial eta-squared = 0.02, on recognition of past counterfactual conditionals. From pre-test to immediate post-test, the direct feedback group improved significantly more than the indirect group: \( F(1, 50) = 4.16, p = 0.047 \), partial eta-squared = 0.08.

4.1.1.2. Results on Production Task

Table 6 displays the means and standard deviations by feedback condition across time on learners’ scores in production of the English past counterfactual conditionals.

Table 6

*Production of Past Conditionals: Mean Test Scores by Feedback Condition across Time*

<table>
<thead>
<tr>
<th>Condition</th>
<th>N</th>
<th>Pretest</th>
<th></th>
<th>Immediate Posttest</th>
<th></th>
<th>Delayed posttest</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mean SD</td>
<td>Mean SD</td>
<td>Mean SD</td>
<td>Mean SD</td>
<td>Mean SD</td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>11</td>
<td>6.36 14.33</td>
<td>36.67 43.77</td>
<td>42.81 45.94</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Indirect</td>
<td>26</td>
<td>9.62 18.43</td>
<td>53.33 38.02</td>
<td>68.08 37.53</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Direct</td>
<td>28</td>
<td>12.76 20.08</td>
<td>83.99 22.84</td>
<td>83.96 26.86</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
A 3 x 3 repeated-measures ANOVA with a one between-subject, one within-subject design was performed on raw percentage scores for production of the past counterfactual conditionals. The between-subject factor was written feedback condition, and the within-subject factor was time. This analysis showed a significant main effect for feedback condition, $F(2, 62) = 9.11, p = 0.00$, partial eta-squared = 0.23, a significant main effect for time, $F(2, 62) = 80.17, p = 0.00$, partial eta-squared = 0.56, and a significant interaction between feedback condition and time, $F(4, 62) = 4.02, p = 0.01$, partial eta-squared = 0.12. The ANOVA results are displayed in Table 7.

Table 7

*Production of Past Conditionals: ANOVA for Feedback Condition by Time*

<table>
<thead>
<tr>
<th>Source</th>
<th>Sums of Squares</th>
<th>df</th>
<th>Mean Squares</th>
<th>$F$</th>
<th>$p$</th>
<th>Partial $\eta^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Feedback</td>
<td>26425.65</td>
<td>2</td>
<td>13212.82</td>
<td>9.11</td>
<td>0.000*</td>
<td>0.23</td>
</tr>
<tr>
<td>Time</td>
<td>99190.84</td>
<td>2</td>
<td>49595.42</td>
<td>80.17</td>
<td>0.000*</td>
<td>0.56</td>
</tr>
<tr>
<td>Time x Feedback</td>
<td>9942.87</td>
<td>4</td>
<td>2485.72</td>
<td>4.02</td>
<td>0.004*</td>
<td>0.12</td>
</tr>
</tbody>
</table>

*p < .05
The significant interaction suggests that the different groups changed differentially from pretest to delayed posttest. This interaction is visually illustrated in Figure 3. The interaction effect has a medium magnitude (partial eta-squared = 0.12) in the greater population.

Figure 3. Mean Scores on Production of Past Conditionals by Feedback Condition
In order to make pair-wise group comparisons (direct vs. control, indirect vs. control, direct vs. indirect) in terms of gain scores in production from pre-test to delayed post-test, separate 2 x 3 repeated-measures ANOVAs with one between-subject (feedback condition), one within-subject (time) design were conducted. The results showed a significant interaction effect when the direct feedback condition was compared with the control and indirect feedback conditions: Direct vs. Control $F(1, 37) = 7.31, p = 0.001$, partial eta-squared = 0.17 (a large effect size), Direct vs. Indirect $F(1, 52) = 4.30, p = 0.016$, partial eta-squared = 0.08 (a medium effect size). When the indirect and the control groups’ gain scores from pretest to delayed posttest were compared, a post-hoc power analysis showed that the observed power for the interaction between feedback condition x time was 0.27, suggesting that the lack of significant interaction may have been due to a small sample size. Post-hoc means comparisons revealed that the production score of the direct group increased significantly more than those of the control ($F(1, 37) = 10.95, p = 0.002$, partial eta-squared = 0.23) and indirect groups ($F(1, 52) = 7.78, p = 0.007$, partial eta-squared = 0.13) from the pre-test to the immediate post-test. Regarding immediate-to-delayed post-test improvement, there were no significant differences between the three groups’ change. The direct feedback group experienced a score growth to a significantly greater extent than the control group did from the pretest to the delayed posttest: $F(1, 37) = 5.67, p = 0.022$, partial eta-squared = 0.13.
4.1.2. Learning of OPREP Types of Relative Clauses

4.1.2.1. Results on Recognition Task

Descriptive statistics on learners’ percentage scores in recognition of the
English OPREP types of relative clauses are provided in Table 8, and graphically
represented in Figure 4.

Table 8

*Interpretation of Relative Clauses: Mean Test Scores by Feedback Condition across Time*

<table>
<thead>
<tr>
<th>Condition</th>
<th>N</th>
<th>Pretest</th>
<th>Immediate Posttest</th>
<th>Delayed Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
</tr>
<tr>
<td>Control</td>
<td>7</td>
<td>57.14</td>
<td>13.80</td>
<td>65.71</td>
</tr>
<tr>
<td>Indirect</td>
<td>20</td>
<td>56.00</td>
<td>25.63</td>
<td>77.00</td>
</tr>
<tr>
<td>Direct</td>
<td>14</td>
<td>47.14</td>
<td>28.94</td>
<td>65.71</td>
</tr>
</tbody>
</table>
Results of the two-way repeated-measures ANOVA showed no significant main effect for feedback condition, $F(2, 38) = 0.57, p = 0.57$, partial eta-squared = 0.03, nor a significant interaction between feedback condition and time, $F(4, 38) = 0.54, p = 0.71$, partial eta-squared = 0.03. The non-significant interaction results
suggest that the different groups’ trajectories over time were not significantly
different. There is, however, a significant main effect for time, $F(2, 38) = 7.94, \ p = 0.00$, partial eta-squared = 0.17 (a large effect size), which indicates that participants
as a whole developed significantly over time in terms of their recognition of the
OPREP types of relative clauses. A post-hoc power analysis revealed that the
observed power for the interaction was 0.17, meaning that the lack of significant
interaction may have been due to a small sample size. These findings are represented
in Table 9.

Table 9

*Interpretation of Relative Clauses: ANOVA for Feedback Condition by Time*

<table>
<thead>
<tr>
<th>Source</th>
<th>Sums of Squares</th>
<th>Df</th>
<th>Mean</th>
<th>$F$</th>
<th>$p$</th>
<th>Partial eta$^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Feedback</td>
<td>1441.51</td>
<td>2</td>
<td>720.76</td>
<td>0.57</td>
<td>0.570</td>
<td>0.03</td>
</tr>
<tr>
<td>Time</td>
<td>8292.25</td>
<td>2</td>
<td>4146.13</td>
<td>7.94</td>
<td>0.001*</td>
<td>0.17</td>
</tr>
<tr>
<td>Time x Feedback</td>
<td>1126.64</td>
<td>4</td>
<td>281.66</td>
<td>0.54</td>
<td>0.707</td>
<td>0.03</td>
</tr>
</tbody>
</table>

* $p < .05$

4.1.2.2. Results on Production Task

Means and standard deviations were computed by feedback condition across
time on learners’ scores in production of the OPREP types of relative clauses in English. These data are presented in Table 10, and Figure 5 is a graphic representation of the findings.

Table 10

Production of Relative Clauses: Mean Test Scores by Feedback Condition across Time

<table>
<thead>
<tr>
<th>Condition</th>
<th>N</th>
<th>Mean Pretest</th>
<th>SD</th>
<th>Mean Posttest</th>
<th>SD</th>
<th>Mean Delayed posttest</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>12</td>
<td>18.75</td>
<td>25.52</td>
<td>34.17</td>
<td>38.72</td>
<td>39.44</td>
<td>40.94</td>
</tr>
<tr>
<td>Indirect</td>
<td>23</td>
<td>24.06</td>
<td>25.99</td>
<td>60.58</td>
<td>36.71</td>
<td>65.00</td>
<td>38.04</td>
</tr>
<tr>
<td>Direct</td>
<td>23</td>
<td>30.65</td>
<td>27.88</td>
<td>72.61</td>
<td>29.85</td>
<td>70.22</td>
<td>33.86</td>
</tr>
</tbody>
</table>
Raw percentage scores on production of the OPREP types of relative clauses were submitted to a 3 x 3 repeated-measures ANOVA using one between-subject, one within-subject design. The between-subject factor was feedback condition (Control, Indirect, Direct), and the within-subject factor was time (pre-test, post-test,
delayed post-test). The repeated-measures ANOVA revealed a significant main
effect for feedback condition, $F(2, 55) = 4.07, p = 0.02$, partial eta-squared = 0.13 (a
medium/large effect size), a significant main effect for time, $F(2, 55) = 32.68, p =
0.00$, partial eta-squared = 0.37 (a large effect size); however, no significant
interaction between feedback condition and time was found. A post-hoc power
analysis revealed that the observed power for the interaction was 0.14. These results
are reported in Table 11.

Table 11

*Production of Relative Clauses: ANOVA for Feedback Condition by Time*

<table>
<thead>
<tr>
<th>Source</th>
<th>$df$</th>
<th>Sums of Squares</th>
<th>Mean Squares</th>
<th>$F$</th>
<th>$p$</th>
<th>Partial $\eta^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Feedback</td>
<td>2</td>
<td>17369.22</td>
<td>8684.61</td>
<td>4.07</td>
<td>0.023*</td>
<td>0.13</td>
</tr>
<tr>
<td>Time</td>
<td>2</td>
<td>37409.22</td>
<td>18704.61</td>
<td>32.68</td>
<td>0.000*</td>
<td>0.37</td>
</tr>
<tr>
<td>Time x Feedback</td>
<td>4</td>
<td>3370.00</td>
<td>842.50</td>
<td>1.47</td>
<td>0.216</td>
<td>0.05</td>
</tr>
</tbody>
</table>

*p < .05

4.1.3. Summary

From all of the analyses conducted for RQ1, it can be concluded that the
different written feedback conditions created for this experiment appear to have a
differential effect on improving learners’ ability to recognize/comprehend and
produce one of the target structures: English past counterfactual conditionals. The
provision of direct written feedback did indeed have a differential effect on learners’
L2 development both in terms of recognition and production of the past
counterfactual conditionals, when compared to the absence of feedback. This effect
was found from pretest to immediate posttest as well as from pretest to delayed
posttest. In terms of immediate-to-delayed posttest improvement, the direct feedback
had a differential effect on learners’ recognition compared to no feedback condition.
Differential effects of different types of written feedback were also found from
pretest to immediate posttest: the direct feedback had a greater effect than the
indirect feedback on improving learners’ ability to recognize and produce the past
conditionals. None of these effects, however, were observed within the current
sample for English OPREP type of relative clauses. This conclusion is discussed in
the next chapter.

4.2. Awareness: Results for the Effects of Type of Feedback on Reported Awareness

Research question 2 asked whether the type of written feedback had an effect
on the level of learners’ reported awareness of the linguistic targets while processing
feedback. For this research question, data came from the participants who had
verbalized during the feedback session. To answer this question, learners were first
regrouped according to the type of reported awareness that they verbalized as their
highest level during the feedback session. As previously explained, four levels of awareness reports, namely [NVR] or [-cognitive change] [-meta-awareness] [-morphological rule], [N] or [+cognitive change] [-meta-awareness] [-morphological rule], [MA] or [+cognitive change] [+meta-awareness] [-morphological rule], [MR] or [+cognitive change] [+meta-awareness] [+morphological rule], were originally identified.

However, since some cells presented expected frequencies lower than 5, the variable of reported awareness was collapsed into two levels. For the past counterfactual conditionals, the concurrent think-aloud protocols were reclassified either [L] or a low level of awareness, or [H] or a high level of awareness. With respect to the OPREP types of relative clauses, the learners’ verbalizations were reclassified either [NVR] or no verbal report, or [N+] or awareness at least at the level of noticing. Final regrouping of learners was performed by considering both the information included in the think-aloud protocols and in the post-exposure questionnaires. Separate chi squares were conducted on frequency counts of each type of reported awareness [L] and [H] for the past conditionals, and [NVR] and [N+] for the relative clauses, respectively, per type of feedback provided.

Table 12 presents the observed and expected counts for level of reported awareness (Low vs. High, No verbal report vs. at least Noticing) by type of feedback according to the targeted linguistic structure. The chi-square analysis revealed a significant effect of feedback type on the type of reported awareness as learners’
highest level on the past counterfactual conditionals, $\chi^2(1) = 3.90, p = 0.048$.

However, no significant effect of feedback type on the learners’ reported awareness at different levels on the OPREP types of relative clauses was found, $\chi^2(1) = 3.21, p = 0.07$.

Table 12

*Chi-square analysis for Type of Feedback by Level of Reported Awareness according to Targeted Linguistic Structure*

<table>
<thead>
<tr>
<th></th>
<th>Reported Awareness</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Low Level of</td>
<td></td>
</tr>
<tr>
<td></td>
<td>High Level of</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Awareness [L]</td>
<td>Awareness [H]</td>
</tr>
<tr>
<td>Past Counterfactual Conditionals</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Indirect</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Actual</td>
<td>9</td>
<td>4</td>
</tr>
<tr>
<td>Expected</td>
<td>(6.3)</td>
<td>(6.7)</td>
</tr>
<tr>
<td>Direct</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Actual</td>
<td>6</td>
<td>12</td>
</tr>
<tr>
<td>Expected</td>
<td>(8.7)</td>
<td>(9.3)</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Actual</td>
<td>Expected</td>
</tr>
<tr>
<td>------------------</td>
<td>--------</td>
<td>----------</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>(15.0)</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>(16.0)</td>
</tr>
</tbody>
</table>

\[ df = 1, \chi^2 = 3.90, p < .05^* \]

Absence of | Presence of
Awareness [NVR] | Awareness [N+]

OPREP Types of Relative Clauses

Indirect

<table>
<thead>
<tr>
<th></th>
<th>Actual</th>
<th>Expected</th>
<th>Expected</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>9</td>
<td>(6.6)</td>
<td>17.0</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>(10.4)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>17</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Direct

<table>
<thead>
<tr>
<th></th>
<th>Actual</th>
<th>Expected</th>
<th>Expected</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>3</td>
<td>(5.4)</td>
<td>14.0</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>(8.6)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>14</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Total

<table>
<thead>
<tr>
<th></th>
<th>Actual</th>
<th>Expected</th>
<th>Expected</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>12</td>
<td>(12.0)</td>
<td>(31.0)</td>
</tr>
<tr>
<td></td>
<td>19</td>
<td>(19.0)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>31</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[ df = 1, \chi^2 = 3.21, p = .07 \]

The significant result found in one of the two chi-square analyses suggests a partial positive answer for RQ 2, that is, the type of feedback did have a differential effect on learners’ levels of reported awareness only when the English past
counterfactual conditionals were targeted. The majority of the participants who received direct feedback reported a high level of awareness while most participants receiving indirect feedback reported a low level of awareness. This significant effect was not found for the English OPREP types of relative clauses.

4.3. Awareness and Learning: Results on the Effects of Reported Awareness

The third research question addressed a relationship between learners’ reported awareness of the linguistic targets at different levels and their development (i.e., the effects of the level of reported awareness on learner’s L2 development in terms of the ability to recognize and produce the target structures). For this research question, data came from the participants who had verbalized during the feedback session and met the selection criteria described in the previous chapter ($n = 29$ for the production of the past counterfactual conditionals, $n = 24$ for the production of the OPREP types of relative clauses). With respect to the recognition test, some cell sizes turned out to be very small (e.g., 6), and the participants’ recognition scores were not included in the analyses for research question 3.

As was the case for research question 2, in order to address the relationship between the levels of reported awareness identified for this study and learners’ improvement over time, participants were redistributed into two groups, [L] and [H], for the past counterfactual conditionals, and [NVR] and [N+] for the OPREP types of relative clauses. Then, separate $2 \times 3$ repeated-measures ANOVAs were performed on learners’ test scores of production of the two target structures. If any
significant interaction effect was found, each ANOVA was followed by means comparisons to analyze exactly where (i.e., pretest to immediate posttest, immediate to delayed posttests, pretest to delayed posttest) the significant effect occurred.

4.3.1. Awareness and Learning: Results for the Past Counterfactual Conditionals

4.3.1.1. Results on Production Task

Means and standard deviations were calculated for levels of reported awareness across time on learners’ scores in production of the English past counterfactual conditionals. These data are shown in Table 13, and are visualized in the plot in Figure 6.

Table 13

Production of Past Conditionals: Mean Test Scores by Reported Awareness across Time

<table>
<thead>
<tr>
<th>Awareness</th>
<th>N</th>
<th>Pretest</th>
<th></th>
<th>Immediate Posttest</th>
<th></th>
<th>Delayed posttest</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>Low</td>
<td>9</td>
<td>17.78</td>
<td>24.38</td>
<td>55.56</td>
<td>34.68</td>
<td>52.22</td>
<td>37.00</td>
</tr>
<tr>
<td>High</td>
<td>17</td>
<td>9.25</td>
<td>17.65</td>
<td>80.20</td>
<td>26.70</td>
<td>88.82</td>
<td>17.64</td>
</tr>
</tbody>
</table>
Percentage scores on production tests of the past counterfactual conditionals were submitted to a 2 x 3 repeated-measures ANOVA with a one between-subject, two within-subject design. The between-subject factor was reported awareness and the within-subject factor was time. This analysis showed a significant interaction.
effect, $F(2, 24) = 5.03, p = 0.010$, partial eta-squared = 0.17 (a large effect size), a significant main effect for reported awareness, $F(1, 24) = 8.16, p = 0.009$, partial eta-squared = 0.25 (a large effect size), and a significant main effect for time, $F(2, 24) = 38.13, p = 0.000$, partial eta-squared = 0.62 (a large effect size). The significant interaction effect indicates that learners who reported a higher level of awareness improved significantly more than those who reported a lower level of awareness in production of the past counterfactual conditionals. These results are summarized in Table 14.

Table 14

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>Sums of Squares</th>
<th>Mean Squares</th>
<th>$F$</th>
<th>$p$</th>
<th>Partial eta$^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Awareness</td>
<td>1</td>
<td>1816.96</td>
<td>1816.96</td>
<td>8.16</td>
<td>0.009*</td>
<td>0.25</td>
</tr>
<tr>
<td>Time</td>
<td>2</td>
<td>48741.17</td>
<td>24370.58</td>
<td>38.13</td>
<td>0.000*</td>
<td>0.62</td>
</tr>
<tr>
<td>Time x Awareness</td>
<td>2</td>
<td>6433.14</td>
<td>3216.57</td>
<td>5.03</td>
<td>0.010*</td>
<td>0.17</td>
</tr>
</tbody>
</table>

* $p < .05$

Post-hoc means comparisons revealed that the production score of the [H] group increased significantly more than that of the [L] group from the pretest to the
immediate posttest, $F(1, 24) = 4.70, p = 0.04$, partial eta-squared = 0.16 (a large effect size). However, no significant difference was found for the two groups’ score changes from the immediate to the delayed posttests, $F(1, 24) = 0.59, p = 0.45$, partial eta-squared = 0.02.

4.3.2. Awareness and Learning: Results for the OPREP Types of Relative Clauses

4.3.2.1. Results on Production Task

Means and standard deviations were calculated for levels of reported awareness across time on learners’ scores in production of the English OPREP types of relative clauses. These data are displayed in Table 15. The plot in Figure 7 visually represents the results.

Table 15

*Production of Relative Clauses: Mean Test Scores by Reported Awareness across Time*

<table>
<thead>
<tr>
<th>Awareness</th>
<th>N</th>
<th>Pretest Mean</th>
<th>Pretest SD</th>
<th>Immediate Posttest Mean</th>
<th>Immediate Posttest SD</th>
<th>Delayed posttest Mean</th>
<th>Delayed posttest SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>[NVR]</td>
<td>9</td>
<td>30.56</td>
<td>26.17</td>
<td>39.26</td>
<td>38.79</td>
<td>52.78</td>
<td>38.33</td>
</tr>
<tr>
<td>[N+]</td>
<td>15</td>
<td>37.00</td>
<td>28.94</td>
<td>87.33</td>
<td>17.92</td>
<td>75.67</td>
<td>22.27</td>
</tr>
</tbody>
</table>
Percentage scores on production of the OPREP types of relative clauses were submitted to a 2 x 3 repeated-measures ANOVA with a one between-subject (reported awareness), two within-subject (time) design. The results showed a
significant main effect for reported awareness, $F(1, 22) = 9.11, p = 0.006$, partial eta-squared $= 0.29$ (a large effect size), a significant main effect for time, $F(2, 22) = 11.74, p = 0.000$, partial eta-squared $= 0.35$ (a large effect size), and a significant interaction effect, $F(2, 22) = 4.30, p = 0.020$, partial eta-squared $= 0.16$ (a large effect size). The significant interaction suggests that learners who reported awareness at least at the level of noticing, or [N+], experienced a significantly higher score growth than those who did not report awareness, or [NVR]) in production of the OPREP types of relative clauses from pretest to delayed posttest. Table 16 summarizes these findings.

Table 16

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>Sums of Squares</th>
<th>Mean Squares</th>
<th>F</th>
<th>p</th>
<th>Partial eta$^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Awareness</td>
<td>1</td>
<td>3745.04</td>
<td>3745.04</td>
<td>9.11</td>
<td>0.006*</td>
<td>0.29</td>
</tr>
<tr>
<td>Time</td>
<td>2</td>
<td>13492.60</td>
<td>6746.30</td>
<td>11.74</td>
<td>0.000*</td>
<td>0.35</td>
</tr>
<tr>
<td>Time x Awareness</td>
<td>2</td>
<td>4945.71</td>
<td>2472.85</td>
<td>4.30</td>
<td>0.020*</td>
<td>0.16</td>
</tr>
</tbody>
</table>

* $p < .05$

Post-hoc means comparisons revealed significant differences between the
score changes made by the [NVR] and the [N+] groups from the pretest to the immediate posttest, $F(1, 22) = 9.27, p = 0.006$, partial eta-squared = 0.30 (a large effect size). However, no significant difference was found for the two groups’ score changes from the immediate to the delayed posttests, $F(1, 22) = 3.93, p = 0.060$, partial eta-squared = 0.15 (a large effect size).

4.3.3. Summary

The findings summarized above suggest a positive answer to research question 3. The level of reported awareness identified through the combination of the think-aloud protocols and a post-exposure questionnaire appears to have a differential effect on the improvement of learners’ ability to produce the targeted linguistic structures over time from pretest to delayed posttest. This relationship was also found from pretest to immediate posttest in terms of the production of the both target structures.

4.4. Reactivity: Results for the Effects of Verbalization Condition

In order to determine whether verbalization condition had a differential effect on learners’ improvement in recognition and production of the two targeted structures, the assessment scores were submitted to separate 2 x 3 repeated-measures ANOVAs, with verbalization condition (+think-aloud vs. –think-aloud) as a between-subjects variable, and time (pretest, immediate posttest, delayed posttest) as a within-subjects variable.
The descriptive data are summarized for verbalization condition across time on learners’ scores in recognition and production according to the type of linguistic targets in Table 17.
### Mean Test Scores by Verbalization Condition across Time

<table>
<thead>
<tr>
<th>Condition</th>
<th>N</th>
<th>Pretest</th>
<th></th>
<th>Immediate Posttest</th>
<th></th>
<th>Delayed posttest</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td><strong>Recognition</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Past Counterfactual Conditionals</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>+Think-aloud</td>
<td>28</td>
<td>49.29</td>
<td>26.80</td>
<td>71.79</td>
<td>32.89</td>
<td>81.79</td>
<td>29.82</td>
</tr>
<tr>
<td>-Think-aloud</td>
<td>33</td>
<td>51.82</td>
<td>27.32</td>
<td>78.48</td>
<td>29.27</td>
<td>85.15</td>
<td>25.02</td>
</tr>
<tr>
<td>OPREP Types of Relative Clauses</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>+Think-aloud</td>
<td>17</td>
<td>58.82</td>
<td>23.95</td>
<td>68.24</td>
<td>28.34</td>
<td>78.82</td>
<td>34.26</td>
</tr>
<tr>
<td>-Think-aloud</td>
<td>24</td>
<td>49.17</td>
<td>25.69</td>
<td>73.33</td>
<td>22.59</td>
<td>74.17</td>
<td>30.35</td>
</tr>
<tr>
<td><strong>Production</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Past Counterfactual Conditionals</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>+Think-aloud</td>
<td>29</td>
<td>12.66</td>
<td>20.61</td>
<td>65.29</td>
<td>35.57</td>
<td>70.00</td>
<td>35.36</td>
</tr>
<tr>
<td>-Think-aloud</td>
<td>36</td>
<td>8.61</td>
<td>16.59</td>
<td>62.45</td>
<td>40.06</td>
<td>71.16</td>
<td>39.43</td>
</tr>
<tr>
<td>OPREP Types of Relative Clauses</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>+Think-aloud</td>
<td>24</td>
<td>34.58</td>
<td>27.53</td>
<td>69.31</td>
<td>35.83</td>
<td>67.08</td>
<td>30.68</td>
</tr>
<tr>
<td>-Think-aloud</td>
<td>34</td>
<td>19.22</td>
<td>24.35</td>
<td>53.24</td>
<td>36.62</td>
<td>58.04</td>
<td>42.85</td>
</tr>
</tbody>
</table>
On the recognition test, the analyses showed no significant interaction effects and no significant main effects for verbalization condition on the two linguistic targets. Significant main effects for time were found on the targeted linguistic structures. The same holds for the production test scores. The non-significant interactions suggest that the concurrent verbalization did not have a significant effect on learners’ L2 development over time. Table 18 exemplifies these findings, as do the plots in Figures 8 through 11.
Table 18

*Reactivity of Concurrent Verbalization: ANOVA for Verbalization Condition by Time*

<table>
<thead>
<tr>
<th>Source</th>
<th>Sums of Squares</th>
<th>df</th>
<th>Mean</th>
<th>F</th>
<th>P</th>
<th>Partial eta²</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Recognition</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Past Counterfactual Conditionals</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TA</td>
<td>267.09</td>
<td>1</td>
<td>267.09</td>
<td>0.48</td>
<td>0.49</td>
<td>0.01</td>
</tr>
<tr>
<td>Time</td>
<td>35491.53</td>
<td>2</td>
<td>17745.77</td>
<td>46.31</td>
<td>0.000*</td>
<td>0.44</td>
</tr>
<tr>
<td>Time x TA</td>
<td>147.27</td>
<td>2</td>
<td>73.63</td>
<td>0.19</td>
<td>0.83</td>
<td>0.00</td>
</tr>
<tr>
<td>OPREP Types of Relative Clauses</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TA</td>
<td>93.91</td>
<td>1</td>
<td>93.91</td>
<td>0.22</td>
<td>0.639</td>
<td>0.01</td>
</tr>
<tr>
<td>Time</td>
<td>10889.83</td>
<td>2</td>
<td>5444.92</td>
<td>10.69</td>
<td>0.000*</td>
<td>0.22</td>
</tr>
<tr>
<td>Time x TA</td>
<td>1120.72</td>
<td>2</td>
<td>560.36</td>
<td>1.10</td>
<td>0.338</td>
<td>0.03</td>
</tr>
<tr>
<td><strong>Production</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Past Counterfactual Conditionals</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TA</td>
<td>58.50</td>
<td>1</td>
<td>58.50</td>
<td>0.10</td>
<td>0.759</td>
<td>0.00</td>
</tr>
<tr>
<td>Time</td>
<td>138598.44</td>
<td>2</td>
<td>69299.22</td>
<td>101.05</td>
<td>0.000*</td>
<td>0.62</td>
</tr>
<tr>
<td>Time x TA</td>
<td>239.03</td>
<td>2</td>
<td>119.52</td>
<td>0.17</td>
<td>0.840</td>
<td>0.00</td>
</tr>
</tbody>
</table>
### OPREP Types of Relative Clauses

<table>
<thead>
<tr>
<th></th>
<th>TA</th>
<th>1</th>
<th>Time</th>
<th>2</th>
<th>0.000*</th>
<th>0.41</th>
</tr>
</thead>
<tbody>
<tr>
<td>TA</td>
<td>2561.85</td>
<td>1</td>
<td>2561.85</td>
<td>3.39</td>
<td>0.071</td>
<td>0.06</td>
</tr>
<tr>
<td>Time x TA</td>
<td>46047.86</td>
<td>2</td>
<td>23023.93</td>
<td>39.12</td>
<td>0.000*</td>
<td>0.41</td>
</tr>
</tbody>
</table>

* *p < .05*

---

128
Table 8. Mean Scores on Recognition of Past Conditionals by Verbalization Condition

<table>
<thead>
<tr>
<th>Time</th>
<th>Pretest</th>
<th>Immediate Posttest</th>
<th>Delayed Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean Score</td>
<td>40.00</td>
<td>50.00</td>
<td>60.00</td>
</tr>
</tbody>
</table>

Figure 8. Mean Scores on Recognition of Past Conditionals by Verbalization Condition
Figure 9. Mean Scores on Interpretation of Relative Clauses by Verbalization Condition

<table>
<thead>
<tr>
<th>Time</th>
<th>Pretest</th>
<th>Immediate Posttest</th>
<th>Delayed Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean Score</td>
<td>50.00</td>
<td>60.00</td>
<td>70.00</td>
</tr>
<tr>
<td>TA Non-think-aloud</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>TA Think-aloud</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
Figure 10. *Mean Scores on Production of Past Conditionals by Verbalization Condition*

![Graph showing mean scores on production of past conditionals by verbalization condition. The x-axis represents time with points at Pretest, Immediate Posttest, and Delayed Posttest. The y-axis represents mean scores with values from 0.00 to 70.00. There are two lines: one for Think-aloud and one for Non-think-aloud. The Think-aloud line shows an increase in mean scores over time, while the Non-think-aloud line shows a slight increase in mean scores over time.]
Figure 11. *Mean Scores on Production of Relative Clauses by Verbalization Condition*

A discussion of the results obtained for each research question is presented in the following chapter.
CHAPTER 5: DISCUSSION AND CONCLUSION

This chapter discusses the results obtained in the present study in the context of the aforementioned research questions. First, the findings on the differential effects of direct and indirect written feedback on L2 development (i.e., Research Question 1) are discussed. This is followed by a discussion on the effect of the type of written feedback on L2 learners’ reported awareness at different levels (i.e., Research Question 2), and the relationship between different levels of reported awareness and L2 development (i.e., Research Questions 3). The reactivity issue of think-aloud protocols is also discussed. Then, the theoretical, pedagogical, and methodological implications of the study are taken into account. The chapter concludes by acknowledging the limitations of the current investigation and discussing suggestions for future research.

5.1. Discussion

5.1.1. Effects of Type of Written Feedback on L2 Learning

The first research question investigated (1) whether the type of written feedback (indirect vs. direct vs. no feedback) impacted L2 learners’ development of the targeted linguistic structures, and (2) if so, whether the effect lasted over one week period. The results of the repeated-measures ANOVAs showed that there is a direct relationship between direct written feedback and development of linguistic accuracy in recognizing and producing one of the two targets: the past
counterfactual conditional in English. The learners who received direct written feedback on the past counterfactual conditional improved significantly more than the control group who did not receive it, both from pretest to immediate posttest and from pretest to delayed posttest. The direct feedback group also showed a greater gain than the indirect feedback group in recognition from pretest to immediate posttest as well as in production from pretest to immediate and delayed posttests. However, the results of this study suggested that the effect of indirect written feedback on a learner’s development of the past counterfactual conditional from pretest to (immediate and delayed) posttests was not significantly different from the effect of the control condition.

The fact that the participants in the direct feedback condition significantly outperformed the participants in the no feedback condition is a finding generally consistent with the results of Bitchener (2008), Bitchener & Knoch (2008), and Sheen (2007). Bitchener (2008) and Bitchener and Knoch (2008) found a positive effect for direct feedback on learners’ improvement from pretest to immediate posttest, when compared to a control group. Likewise, Sheen (2007) showed that a direct-only feedback group outperformed a control group on immediate posttests.

However, the control group in Bitchener (2008) received a brief, general comment, and Sheen’s (2007) control group completed only pretest and posttests, and not any treatment tasks; thus, care needs to be taken in comparing these findings. In this study, the only difference between the direct and the control groups was the
presence/absence of direct written feedback, and it was possible to show that the provision of direct feedback was effective in resulting in improved accuracy of the English past counterfactual conditional. Moreover, direct feedback proved to be effective in Korean EFL learners’ development of the past counterfactual conditional in both the short and longer (1 week later) terms. This is a meaningful finding given the finding from Sheen (2007) that only direct metalinguistic correction was effective in improving learners’ accuracy on the delayed posttests, while direct-only correction was not.

The fact that there was no significant difference found between the indirect feedback and the control groups in this study is inconsistent with the findings of several other studies (e.g., Ashwell, 2000; Fathman & Whalley, 1990; Ferris & Roberts, 2001). A major difference across the studies that may account for the contrast in findings is the type of assessment measures used. Whereas the present study used different versions of the recognition and production (written story-retelling task) tests to measure L2 learning, other studies used revision tasks of the same essay. The treatment task used in Fathman and Whalley (1990) is similar to that of this study: they asked participants to write a story about a sequence of eight pictures during 30 minutes, and provided indirect feedback on grammatical errors by means of underlining. However, Fathman and Whalley (1990) used the rewrites as the measure of improvement, and found a significant difference between those who received indirect feedback and those who did not. Ashwell (2000) provided Japanese
EFL learners with form feedback consisting of underlining or circling grammatical, lexical, and mechanical errors or of using cursors (\(^\wedge\)), although content feedback was also given on one of two writing versions of a composition. He found that feedback groups receiving form feedback made significantly more improvement than no feedback group on revisions of the same composition. Ferris and Roberts (2001) also showed that an indirect feedback group who received feedback involving errors underlined significantly outperformed the control group on a self-editing task. Given that revision studies cannot provide direct evidence for the effects of written feedback on L2 development or learning and given the differences in the above-mentioned studies, it remains unsupported whether indirect feedback involving grammatical errors underlined facilitates L2 development. Nonetheless, considering that there was low observed power when the indirect and the control groups’ gain scores over time (from pretest to delayed posttest) were analyzed in this study together with the changes in the actual mean scores for two groups illustrated in the plots in Figures 1 and 2, it seems conceivable that any effect of indirect feedback on L2 development over time was concealed in the sample of this study.

In contrast to the finding of Lalande (1982), where indirect feedback was shown to be more effective than direct feedback, this study found that the direct feedback group learners improved significantly more than the indirect feedback group from pretest to immediate posttest both in terms of recognition and production of the past counterfactual conditional. The plausible explanations for this
discrepancy seem to be related to the type of indirect feedback used and the presence/absence of rewrite tasks during the treatment. In terms of type of indirect feedback, the present study operationalized indirect feedback as indication of linguistic error through underlining and/or cursors (∧), whereas Lalande (1982) provided coded feedback through a systematic marking, which is relatively more specific and explicit than the one used in this study. Unlike the current study, Lalande (1982) asked all group learners to rewrite the same essay after providing feedback. These differences in design could have impacted the dissimilar results.

An additional goal of the present study was to explore the possible links between type of written feedback, linguistic targets, and L2 learning. The significant relationship between direct written feedback and L2 development, which was found for the past counterfactual conditionals, was not shown when feedback was provided on the OPREP types of relative clauses in English. The relationship between indirect written feedback and L2 development was not shown to be significant either. The smaller sizes of the sample than the ones considered for the past counterfactual conditionals as well as very low statistical power obtained for the interaction of feedback condition and time on the learners’ scores for the English OPREP relatives suggest that any effect of feedback condition on L2 development over time might have been hidden in the sample of this study for the English OPREP relatives. However, it is noteworthy that large effect sizes were found for the past counterfactual conditionals while small effect sizes were found for the
English OPREP relatives. The differences in effect sizes support that the effectiveness of written feedback may vary depending on the type of linguistic targets. The discrepancy in the results between the two linguistic targets may be explained in terms of the salience of the targets and the amount of exposure to feedback, which is discussed in relation to the learners’ awareness in the following section.

A common finding, observed across the type of linguistic targets, is that the participants in this study, regardless of feedback condition, significantly improved from pretest to delayed posttest. This suggests that participating in three written story-retelling tasks and three feedback sessions were effective in the Korean EFL context. Although the control group learners did not receive any feedback on the targeted linguistic structures, they did look at their own writings during the feedback session. Given that data from those who reported on the post-exposure questionnaires that they had consulted with any external source (e.g., textbooks) about the targets during the period of experiment were discarded from the final participant pool, the chances are very low that the control group’s improvement over time was due to the aid of external resources. Responses from participants in the control group on the debriefing questionnaires revealed that they thought the purpose of the overall experimental tasks was to make them improve by engaging in self-correction. It seems, thus, that reviewing their own output, noticing the gap, and trying to figure out (possible) problems themselves had an effect on learners’ L2
development. Since the targeted structures were not completely new to the participants, it may have been possible that even those in the control group could restructure their knowledge about the targets.

5.1.2. Effects of Type of Written Feedback on Reported Awareness

The second research question concerned the relationship between type of written feedback and learners’ reported awareness of the linguistic targets at different levels while interacting with feedback. Results showed that type of written feedback was related to the level of reported awareness (reduced to low [L] and high [H]) when the English past counterfactual conditionals were targeted. However, the type of written feedback did not appear to have a significant effect on learners’ reported awareness (reduced to [NVR] and [N+]) when the English OPREP types of relative clauses were targeted.

The results, presented in Table 14, show that the learners who received direct feedback on the past counterfactual conditional tended to report awareness at higher levels whereas receiving indirect feedback in many cases did not lead learners to report awareness of the targeted structure at higher levels in this study. About 67 percent of the learners who received direct feedback and provided evidence of registration of the target structure with awareness reported awareness at the level of reporting or understanding while about 33 percent of those in the same condition reported awareness at the level of noticing. However, the difference was
reversed for the indirect group learners. Seventy percent of the learners who received indirect feedback and registered the target structure with awareness reported awareness at the level of noticing while 30 percent of those in the same condition reported awareness at the level of reporting or understanding. This difference in pattern suggests that the different types of written feedback did promote learners’ awareness of the past counterfactual conditional at different levels in the current empirical study. This finding is largely in line with the findings of Rosa and Leow (2004a), who showed a clear association of more explicit information on their target structure with higher levels of reported awareness in the context of a computerized problem-solving task.

With respect to the English OPREP types of relative clauses, learners’ attentional processes were not shown to be significantly affected by type of written feedback. Table 14 indicates that the learners who received direct feedback on the OPREP relatives did tend to report awareness at least at the level of noticing. However, in the indirect feedback condition, roughly half of the learners reported awareness of the targeted structure at least at the level of noticing while the other half did not report awareness at any level. These results seem to suggest that direct written feedback was effective in drawing learners’ attention to the OPREP relatives, and indirect feedback was not as successful as direct feedback (8 out of 17 learners in the indirect feedback group vs. 11 out of 14 learners in the direct feedback group) in leading learners to report registration of the target with awareness. For the chi
square analysis on the OPREP relatives, the original four levels of reported awareness were reduced to the two categories, namely, no verbal report [NVR] and awareness at least at the level of noticing [N+], because some cells presented expected frequencies lower than 5. Had more learners in the indirect feedback + think-aloud group reported awareness at the level of noticing or at higher levels, it could have been possible to examine whether type of written feedback and reported awareness of the OPREP relatives at the low and high levels of awareness were significantly related. Unfortunately, this more fine-tuned analysis was not possible in this study, and it is difficult to make a direct comparison with the results on the past counterfactual conditional.

When the OPREP relatives were targeted, 12 out of 31 (39%) learners who received written feedback and had a think aloud failed to report awareness of the target structure whereas only 3 out of 34 (0.09%) learners who received feedback and did a think-aloud did so when the past counterfactual conditionals were targeted. This is a notable difference in frequency of no verbal reports of the target structures. This inconsistency in results may be attributed to the amount of exposure to feedback and the type of targeted linguistic structures. An examination of the learners’ writing productions during the treatment sessions as well as their verbal reports revealed that the learners in many cases did not receive the same amount of feedback on the two target structures, and they seemed to have not paid an equal degree of attention to the two targets during the treatment. In this study, written
feedback was provided only when learners made errors. The learners, overall, made more frequent errors in using the past counterfactual conditionals compared to the OPREP relatives, and therefore received a higher amount of written feedback on the past conditionals than on the OPREP relatives. This difference in the amount of exposure to feedback might have impacted the results by making the learners more attuned to the past counterfactual conditional than to the OPREP relatives.

Another possible explanation for the noticeable difference in frequency of noverbal reports of the two target structures relates to the type of linguistic target. One of the purposes of including two targeted structures that seem to vary in salience was to explore whether the type of linguistic target may affect the extent to which written feedback influence learners’ attentional processes and subsequent L2 learning. The past counterfactual conditional in English is possibly perceived to be more salient than the OPREP types of relative clauses possibly due to, for instance, a contrast between two words (e.g., “who,” “for”) and six words (e.g., “if”, “had studied”, “would have passed”) and a difference in scope (e.g., If I had studied hard vs. The person who she was looking for is tall). Given the think-aloud data, in which the learners receiving indirect written feedback tended to report awareness of the past counterfactual conditional (at any level) more frequently than they reported awareness of the OPREP relatives, it seems that the past counterfactual conditionals appeal more strongly to the learners than the OPREP relatives. In addition, the learners were exposed to feedback on the two structures at the same time in each
composition. Thus, it seems plausible that the learners’ attention was directed to more salient L2 forms in many cases. In fact, those who did not demonstrate awareness of the OPREP relatives while thinking-aloud reported on the debriefing questionnaire that they had noticed the past counterfactual conditionals during feedback session, but they did not mention about the OPREP relatives. These results are largely in line with the findings of several other studies (e.g., Leow et al., 2003; Mackey, 2006). As mentioned earlier, Mackey (2006) found that the degree of L2 learners’ noticing of oral feedback differed depending on the type of linguistic item. Likewise, Leow et al. (2003) found that the learners exposed to perceptually more salient forms (Spanish present perfect) reported more noticing than those exposed to less salient forms (Spanish present subjunctive).

5.1.3. Link between Reported Awareness and L2 Learning

The third research question asked whether there is a relationship between learners’ reported awareness of the linguistic targets at different levels and their development. The results of the present study provided a positive answer to the question. Different levels of reported awareness were shown to have differential effects on learners’ L2 development in terms of the ability to produce the targeted linguistic structures.

The learners who reported awareness of the past counterfactual conditionals at the level of understanding (a higher level of awareness) improved significantly
more than those who reported awareness of the targeted structure at the level of noticing (a lower level of awareness) on the production tests. The level of reported awareness significantly affected the learners’ improvement of the past counterfactual conditionals over time both from pretest to immediate posttest and from pretest to delayed posttest.

With respect to the OPREP types of relative clauses, the learners who reported awareness of the OPREP relatives at least at the level of noticing showed a significantly greater gain in production than those who gave no evidence of awareness of the targeted structure in their reports both from pretest to immediate posttest and from pretest to delayed posttest. It was not possible to explore whether the learners who reported awareness of the OPREP relatives at the level of understanding improved significantly more than those who reported awareness of the target at the level of noticing because only 15 out of 24 participants reported awareness of the target structure.

However, these results together seem to suggest that there is a significant association of higher levels of reported awareness with learners’ development of the ability to use the targeted linguistic structures. The results also indicate that the differential effects of reported awareness were maintained over a week period. These results concur with those of previous studies (e.g., Leow, 1997; Rosa & Leow, 2004a; Rosa & O’Neil, 1999; Sachs & Suh, 2007) showing a significant relationship between higher levels of reported awareness of a targeted form and better L2
outcomes.

5.1.4. Reactivity of Online Verbal Reports

The fourth research question addressed whether the verbalization condition (think-aloud vs. non-think-aloud) during the feedback session had an effect on learners’ development of linguistic targets. Results showed that thinking aloud did not significantly affect learners’ L2 learning although the written story-retelling task and the feedback session improved learners’ recognition and production of the English past counterfactual conditional and the English OPREP relatives.

The finding that thinking aloud was not reactive in terms of accuracy coincides with those of Bowles (2008), Bowles and Leow (2005), Leow and Morgan-Short (2004), Sachs and Polio (2007, Study 2), Sachs and Suh (2007), and Sanz et al. (2008, Exp.1). These studies all found, as did this study, no significant effect of concurrent verbal reports on learners’ posttest performance or their improvement from pre- to posttests when comparing learners who were told to speak aloud whatever thoughts that ran through their minds and those in a silent control condition.

At the same time, the current finding contrasts with the results of Sachs and Polio (2007, Study 1) and Sanz et al. (2008, Exp.2). The discrepancy could be justified in light of the language of verbalization and the different nature of the treatment tasks used in these investigations. In terms of the language of verbalization,
Sachs and Polio (2007) asked learners to think aloud in their L2, whereas the participants in this study verbalized their thoughts in their L1 (except when they read aloud their L2 output or feedback). Thinking aloud in the L2 may have affected learners’ internal processes, given that producing the L2 pushes learners to engage in syntactic processing required for production (Swain, 1995), drawing attention to the task of verbalization rather than to the treatment task.

Inconsistent findings between the present study and Sanz et al. (2008, Exp.2) may be due to the different type of treatment tasks employed in each of the studies. Whereas learners in this study were asked to think aloud while reading aloud their writings with feedback provided, learners in Sanz et al.’s (2008) second experiment verbalized while performing input-decoding practice with explicit feedback. Unlike this study, Sanz et al. (2008, Exp.2) did not use a type of reading task, and provided metalinguistic feedback. These differences could have influenced the results of the studies.

5.2. Conclusion

5.2.1. Implications for Theory, Pedagogy, and Methodology

The findings obtained in the present study have theoretical, pedagogical, and methodological implications. On the theoretical side, this study contributes to the discussion on whether and how providing written feedback varying in their degree of directness on learners’ written production may affect the development of accuracy
in using two types of linguistic targets. This study has shown that providing direct feedback on a more salient linguistic structure can enhance learners’ ability to acquire and use the targeted structure in both short and longer (1 week later) terms. In addition, the study provided additional empirical support for the association of higher levels of awareness with better L2 learning outcomes (e.g., Leow, 1997; Leow, 2001a; Rosa & Leow, 2004a; Rosa & O’Neill, 1999), and to Schmidt’s (1990, 1993, 1995) arguments concerning the cognitive processes associated with different levels of awareness.

From a pedagogical perspective, the current study has provided support for the merit of providing direct feedback on a more salient linguistic target (the past counterfactual conditionals in this study) in L2 learners’ written production as an effective instructional method, as evidenced by the significant differences found between control and direct feedback learners as well as between indirect and direct feedback learners. In the context of written story-retelling tasks, direct written feedback could draw learners’ attention to their incorrect use of the targeted structure and, furthermore, facilitate learners to develop higher levels of awareness and develop their ability to use it accurately.

The present empirical study also has some methodological implications. This study addressed a methodological concern of whether concurrent verbal reports (i.e., think-aloud protocols) can be employed as a research tool to measure learners’ internal processes without altering the very constructs they set out to reflect. The
results of this study showed that thinking-aloud while reading aloud learners’ own production with written feedback provided did not significantly affect L2 learning. However, caution is still called for in interpreting the results of nonreactivity in terms of accuracy, because we cannot assure whether the internal processes were indeed unchanged or whether the “end state” (Ericsson & Simon, 1993) as measured by learning outcomes was not different but the processes altered.

In addition, the study contributed to the SLA methodology by fine-tuning the coding of awareness levels. Leow (1997) coded learners’ think-aloud protocols according to three criteria to distinguish different levels of awareness. In his coding scheme, verbal reports were coded as the second category, [+cognitive change (+CC), +meta-awareness (+MA), -morphological rule (-MR)], when learners self-reported subjective experiences, but did not verbalize any underlying rule. In the third category, [+cognitive change (+CC), +meta-awareness (+MA), +morphological rule (+MR)], learners demonstrated a report of subjective experiences and verbalized underlying rules metalinguistically. In this study, when participants in the indirect feedback group showed a sign of subjective experiences such as “ah!” while interacting with the feedback provided, and continued to make (and verbalized) self-corrections for the remaining part, these were coded as meta-awareness ([MA]), even if the participants did not make an explicit report of subjective experiences such as “so I have to remember that” (Leow, 1997, p.479). In line with Schmidt’s (1995) two levels of awareness, these were considered to
represent awareness at the level of understanding, even if the participants did not verbalize the underlying rule as in Leow (1997). In other words, by taking into account the learners’ subsequent language behavior throughout the treatment task performance, this study found that awareness at the level of “reporting” in Leow’s (1997) coding scheme can, in fact, include two levels of awareness, “reporting” and “understanding.” Furthermore, these results showed that there can exist different levels of processing at each level of awareness.

5.2.2. Limitations and Suggestions for Future Research

There are inevitably limitations to the present study which should be acknowledged and considered when future research is designed and conducted. A pair of limitations relate to the pool of participants used in this study. The first limitation concerns experimental mortality and cell size. The original participant pool of 102 was randomly assigned to one of five experimental conditions, and began the current study. Differential participant attrition and elimination across conditions resulted in uneven cell distribution. Fewer participants were assigned to the control group when compared to the other feedback groups (e.g., \( n = 11 \) versus \( n = 26 \) and \( n = 28 \) for production of the English past counterfactual conditionals). The uneven mortality might have affected the results of this study. However, although the uneven cell size is not ideal, it is not deemed that the differential mortality caused a bias for one group over the other since the three groups’ mean pretest
scores were not statistically significantly different from each other. Nonetheless, more even distribution of cells could strengthen the results.

The second limitation deals with sample size and power. The final sample size used in the present study (N = 41 to 65) was somewhat small after eliminating participants who did not fulfill the selection criteria. In some cases, in order to preserve the sample size, it was necessary to raise the maximum cut-off scores required on the pretests for a participant to be included in the final analyses (especially for recognition test scores). It seems that this rather small sample size has caused the low statistical power, which is determined by sample size and effect size, in some of the analyses conducted in this study. This may have influenced the statistical results and might have caused the lack of statistical significant differences in some analyses of this study. A larger final participant pool with lower levels of prior knowledge of the targets could increase the strength of the findings.

In a related shortcoming, it would have been ideal to run three-way repeated-measures ANOVAs on learners’ test scores, with feedback condition as the between-subject factor, type of targets (English past counterfactual conditional vs. OREP types of relative clauses) and time (pretest vs. immediate posttest vs. delayed posttest) as the within-group subject factors in order to more precisely investigate the possible interaction of the type of linguistic target with feedback condition and time on L2 learning. This was not possible in this study since the sample size of the control group was extremely small after screening outliers. Future research is
expected to address the role of type of linguistic target as a separate independent variable by collecting data from a larger number of participants with comparable amounts of prior knowledge across two target structures at the outset of the study.

The fourth limitation relates to the feedback targeting the OPREP types of relative clauses. There is a distinction between a pied-piping (e.g., *the person with whom I was talking*) and a preposition stranding (e.g., *the person who(m) I was talking with*) depending on where the preposition is placed in the OPREP relatives. As in other studies on English relativization (e.g., Doughty, 1988, 1991; Izumi, 2000, 2002), the present study used the preposition stranding construction and *who* instead of *whom* in the feedback provided and all the test sentences, although the use of *whom* by a few participants in their stories was not considered incorrect. Given that the context of story-retelling task is much less formal than, for example, that of formal essay writing, it was deemed appropriate to keep participants from solely relying on the prescriptive rule of using *whom* where *who* can be widely used in American English. However, criticism may be raised against the use of *who* in place of *whom*, since *whom* is the correct form of objective relative pronoun from the perspective of prescriptive grammar.

The fifth limitation pertains to the short time which elapsed between the immediate and delayed posttests. Future studies could examine the issues of written feedback and learner awareness through their longer-term effects (e.g., Bitchener, 2008) while attempting to control possible effects of external variables.
The sixth limitation is rooted in the lack of baseline data collection. Although the written story-retelling task was designed carefully based on the consultation with native English-speaking SLA researchers, this study did not obtain baseline data from native-speakers of English. Baseline data would allow ensuring that native-speakers of English actually use the targeted structures doing the same task. The collection of baseline data therefore could provide the stronger grounds for providing feedback in the target context which was created in this study and increase the robustness of the research design.

Another limitation of the study is the fact that the sample size used in the pilot study was small. In an effort to make sure that the different versions of the story-retelling task were comparable in terms of task difficulty, pilot testing was carried out and it was possible to ensure the comparability of the six versions. A larger-scale pilot study could strengthen the validity of this study.

Finally, the findings of this study should be interpreted with caution and applied keeping in mind the context and learner population included in the present study. All participants were L1 Korean university EFL learners, and it was possible to enhance the internal validity of the study by situating the current investigation in a foreign language context where the participants had a similar type and amount of exposure to the target language. However, it is an empirical question whether the same results would hold for, for example, learners in second language or heritage language learning contexts where learning usually takes place in more naturalistic
environment and through less instruction compared to foreign language context. Likewise, the results of this study may not be generalizable to learners from different L1 backgrounds. Future research could additionally address the role of written feedback in L2 development and relative efficacy of different types of written feedback including different linguistic targets (e.g., morphology, other syntactic structures), different L2s other than English, varied proficiency level of participants, and other feedback types (e.g., coded feedback). It would also be necessary to explore whether similar or different effects on L2 development can be obtained using different genre tasks and less focused approach (i.e., treating a broader range of linguistic errors) before generalizations can be drawn.

5.2.3. Conclusion

Some limitations notwithstanding, the present empirical investigation extended our understanding of type of written feedback, awareness, linguistic targets, and L2 development. This study sought to investigate the effectiveness of differential written feedback for L2 learning in relation to participants’ attentional processes during the interaction with the feedback. It also aimed to elucidate the possible links between type of feedback, linguistic targets, learner awareness, and L2 development, and test the reactive effects of concurrent verbal protocols on L2 development.

First, this study found that providing direct written feedback on a more
salient structure (English past counterfactual conditionals) in a L2 writing task leads to development on both immediate and delayed posttests, while indirect written feedback might not have developmental benefits when compared to a no feedback condition. However, neither direct nor indirect written feedback provided on a less salient structure (OPREP types of relative clauses) was shown to have a significant effect on learners’ development when compared to a no feedback condition.

Second, the study found that the provision of direct written feedback leads to a higher level of reported awareness of the more salient structure (English past counterfactual conditionals) while providing indirect written feedback leads to a lower level of reported awareness of the structure. However, type of written feedback did not appear to have a significant effect on learners’ reported awareness of the less salient structure (OPREP types of relative clauses) at different levels (no verbal report of awareness versus. awareness at least at the level of noticing). Although a trend indicated an association of type of written feedback with different levels of reported awareness of the less salient structure (OPREP types of relative clauses), this did not reach a statistical significance.

Next, the present study found a significant relationship between learners’ reported awareness of the targeted structures at different levels and their development. Higher levels of reported awareness were significantly more effective than lower levels of reported awareness in promoting the improvement of learners’ ability to use the two targeted linguistic structures. This relationship was found not
only from pretest to immediate posttest, but also from pretest to delayed posttest.

Finally, in this study, thinking-aloud was not found to be reactive in terms of accuracy. Thinking aloud during the feedback session did not significantly impact learners’ L2 development both in terms of recognition/comprehension and production of the two targeted linguistic structures.

In summary, the current study shed some light on the links between type of written feedback, learner awareness, type of linguistic structure, and L2 learning, and contributed to the body of reactivity research in the field of SLA.
APENDIX A
Background Questionnaire

당신의 이름은 무엇입니까? __________________

당신의 나이는 몇 살입니까? ______________성별을 적어 주세요. __________

당신의 모국어는 무엇입니까?  ____________________

영어권 국가에 거주한 적이 있으십니까? ___________________________

만일 위의 질문에 예라고 답하셨으면, 국가와 기간을 적어주세요.
__________________________________________________________________

모국어 이외에 할 줄 아는 다른 언어가 있으십니까? 있다면, 어떤 언어인지 적어주세요. __________________

만일 위의 질문에 예라고 답하셨으면, 그 언어(들)을 얼마나 잘 아시는지 적어 주세요. (예, 유창하게 할 줄 안다, 읽을 줄만 안다, 약간의 회화 능력, 등)
__________________________________________________________________

얼마나 오랫동안 영어를 공부하셨습니까? ______________________________

영어를 어디서 배웠는지 (배우고 있는지) 적어 주세요. (예, 고등학교에서, 사설 영어 학원에서, 등)
__________________________________________________________________

왜 영어를 배우십니까? (예, 필수/의무 과목이라서, 개인적으로 관심이 있어서, 등)
__________________________________________________________________

교실 밖에서(수업 시간 외에) 영어를 사용하는 시간이 일주일에 얼마나 됩니까?______________________________
당신의 전공이 무엇입니까? ______________________________________
What is your name?
______________________________________________________

How old are you? _______________ Are you male or female? ________________

What is your first language?
_______________________________________________

Have you ever lived in an English-speaking country? ____________

If so, in which country have you lived? _______________ Length of time? ______

Do you know any other languages?
_________________________________________

If so, how well do you know them? (For example: fluently, only reading, some conversational ability, etc.)

—

How long have you been studying English?
___________________________________

Please explain where you have studied English. (For example: in high school, in a private language school, etc.)

—

Why are you studying English? (as a requirement, personal interest, etc.)

—

How many hours a week do you use English outside the classroom?
_______________

What is your major area of study?
___________________________________________
APENDIX B

Debriefing Questionnaire: Version A for the Think-aloud Group

다음의 질문들을 주의 깊게 읽은 후 가능한 자세하게 답변해 주세요. 그러나 절대로 질문을 미리 읽어서는 안 됩니다. 반드시 순서대로 답변해 주시고 한 번 지나간 문항으로 다시 돌아가서 답변을 바꿀 수는 없습니다. (Please read the following questions carefully and answer them in as much detail as possible. Please do not read ahead. Please answer the questions in order and do not go back.)

1. 그림카드를 이용해 영어로 이야기를 만들어 보고 (자신이 쓴 글을 다시 읽어보는) 피드백 세션을 가졌습니다. 전체적으로 이 활동들의 목적이 무엇이었다고 생각합니까? (여러분이 생각/접작하는 이 activities의 구체적인 이유나 목적이 있다면, 그것을 적어 주면 됩니다.)

2. 한글로 읽은 이야기를 영어로 작문했을 때 어떤 어려움이나 문제점을 느꼈습니다 안가요?

3. 여러분이 만든 영어 이야기를 다시 읽어보는 피드백 세션에서, 선생님이 특정한 무언가(어떤 내용이나 항목)에 초점을 맞추고 있었나요 느낌치가 생각해줍니까?

4. 영어 지문을 소리내어 읽는 것에 대한 생각이나 느낌을 적어주세요.

5. 영어 지문을 읽는것과 중간중간에 떠오르는 생각들을 소리내어 말하는 것에 대한 생각이나 느낌을 적어주세요. (eg., 신선했다, 재미있었다, 어색하게 느껴졌다, 글의 내용을 이해하는데 방해가 되는 듯 했다, 등등)

6. 영어로 이야기 만들기와 피드백 세션을 통해 (각각 또는 전체적으로) 무엇가 배운 게 있습니까? 영어와 관련해 어떤 홍미로운 점 또는 여러분 검을 발견하셨습니까? (예, 동사, 전치사, 단어, 등등)
7. 만약 여러분이, 영어로 이미 지나간 과거의 어떤 일이나 사실에 대해 후회하다거나 그 반대 상황을 상정해서 이야기하는 법을 친구에게 설명해 주어야 한다면, 그 친구에게 도움이 될 만한 일종의 "규칙"같은 것을 말해줄 수 있겠습니까? 만약 그렇다면, 그 "규칙"을 적어 주세요.

8. 만약 여러분이, 영어로 두 인물이나 사물의 관계를 한 문장으로 설명하는 법을 친구에게 설명해 주어야 한다면, 그 친구에게 도움이 될 만한 일종의 "규칙"같은 것을 말해줄 수 있겠습니까? 만약 그렇다면, 그 "규칙"을 적어 주세요.
Debriefing Questionnaire: Version B for the Non-think-aloud and Control groups

다음의 질문들을 주의깊게 읽은 후 가능한 자세하게 답변해 주세요. 그러나 절대로 질문을 미리 읽어서는 안 됩니다. 반드시 순서대로 답해 주시고 한 번 지나간 문항으로 다시 돌아가서 답변을 바꿀 수는 없습니다. (Please read the following questions carefully and answer them in as much detail as possible. Please do not read ahead. Please answer the questions in order and do not go back.)

1. 그림카드를 이용해 영어로 이야기를 만들어 보고 (자신이 쓴 글을 다시 읽어보는) 피드백 세션을 가졌습니다. 전체적으로 이 활동들의 목적이 무엇이었다고 생각합니까? (여러분이 생각/감작하는 이 activities의 구체적인 이유나 목적이 있다면, 그것을 적어 주면 됩니다.)

2. 한글로 읽은 이야기를 영어로 작문했을 때 어떤 어려움이나 문제점을 느껴셨습니까?

3. 여러분이 만든 영어 이야기를 다시 읽어보는 피드백 세션에서, 선생님이 특정한 무언가(어떤 내용이나 항목)에 초점을 맞추고 있었다고 느껴거나 생각했습니까?

4. 영어로 이야기 만들기와 피드백 세션을 통해 (각각 또는 전체적으로) 무언가 배운 게 있습니까? 영어와 관련해 어떤 흥미로운 점 또는 어려운 점을 발견하셨습니까? (예, 동사, 전치사, 단어, 둘등)

5. 만약 여러분이, 영어로 이미 지나간 과거의 어떤 일이나 사실에 대해 후회한다거나 그 반대 상황을 상정해서 이야기하는 법을 친구에게 설명해 주어야 한다면, 그 친구에게 도움이 될 만한 일중의 “규칙”같은 것을 말해줄 수 있겠습니까? 만약 그렇다면, 그 “규칙”을 적어 주세요.

6. 만약 여러분이, 영어로 두 인물이나 사물의 관계를 한 문장으로 설명하는 법을 친구에게 설명해 주어야 한다면, 그 친구에게 도움이 될 만한 일중의 “규칙”같은 것을 말해줄 수 있겠습니까? 만약 그렇다면, 그 “규칙”을 적어 주세요.
APPENDIX C

Picture Prompts

Picture A1

Min-Woo
walk street
last Thursday

find wallet bills
identification cards

Picture A2

little while think
seem like fate
send
Picture A3

what right

situation

know

Picture A4

decide find not find
owner phone number police
wallet

take
Luckily owner Ji-Hyun

police

trace back

pleasantly surprised

see wallet money again
part which worry about most financial status

cause many problems

if not get back money
find herself a lot of trouble if not be able to pay bills on time

be obliged to if not send in payments pay late fees on time
if not get back replacement drivers’ license

if Min-Woo not return wallet

if Min-Woo not return wallet

report loss

residence card
offer

reward

become friends

Picture A13

personal characteristic

which Min-Woo be proud of

honesty

Picture A14
Sun-Young, a co-worker who works with Ji-Hyun, shares problems with Ji-Hyun.

Mi-Na, a friend who works with Ji-Hyun, shares problems with Ji-Hyun.

A police officer handles a situation.

Sun-Young says, "Honesty and quality, which I learn from."
Hee-Won

junior university

Seoul

look back freshman year

situation that regrets about
man who call Min-Ki

call Hee-Won

go out with

graduate from same high school which Hee-Won graduate from
take courses
which she enroll in

night before final exam economics

ask go see movie
Picture B7

like course
that much

feel like study

Picture B8

think fun
go out with
decide movie

person who with Min-Ki
be fond of
next day  fail  have to  
test  repeat  
course

regret  choice  if  stay  
think  go out  library  
see  movie
naturally

Picture B11

study

more

exam

if

stay

library

Picture B12

if

be prepared

test

get

good grade
instead
have to
repeat
course summer

feel bad because
need to ⇐ if
repeat pass
course summer
Picture B15

if take course

summer

do things

much more fun

Min-Ki
woman who work advertising agency

Mike’s wife Lisa important businessperson

Mike married to
Picture C3
company which work for
large

Picture C4
Lisa set 5:00am
because meeting which participate in 6:30am
Lisa

Mike

go to bed

husband

come into

change

alarm clock setting

6:00am
Lisa

wake up

6:00am

oversleep

run

bus

miss
30 minutes

late

meeting
unfortunately

not first time
late

inability

on time

something which

boss

Gary

upset about
excuse which
she rely on
not convincing
Gary

situation prevent
if couple talk more
if talk about work schedules

Mike not change alarm clock setting

if Lisa not oversleep

not miss bus
Picture C15

if catch bus → not late meeting

Picture C16

if Lisa arrive on time

if Gary not so disappointed again
Picture D1

man       who         Amy
engage  to                              go    out
suggest                  celebrate
                        anniversary

Picture D2

fiancé       Steve
work
company
which       Amy
work
for
Steve is a person who easily talks to Amy. They talk about topics which are not limited to work.
evening

go downtown see poster concert which they look forward to

last day performance miss
Picture D7

decide  go
concert  car
late

Picture D8

way  theater
got stuck  traffic
Picture D9

arrive theater
find

tickets sold out

Picture D10

think

if plan activities in advance → problem avoid
Picture D11

if decide leave

avoid

leave earlier

heavy traffic

Picture D12

if not get caught

buy tickets
in fact
if
make a reservation
phone
attend concert for sure

in the end
not
have

very good time
night
if attend concert 

better anniversary celebration both of them
last summer vacation
Hee-jin                   assistant
English camp                 Jeju-Island

yesterday Hee-Jin
meet                               friends
who                        she
work                          with
camp
David (friend) who she talk with various things

not good-looking he have admirable qualities which Hee-Jin look up to
Picture E5

friend       Susan

have         beautiful          appearance
which        Hee-Jin

dream of

Picture E6

aspect       personality       seriousness
work
which        Hee-Jin

learn from
talk other opportunities that take advantage of last summer

if Hee-Jin take Statistics
stay Seoul summer session
if she enroll in driving school → take driving test

if David work full-time museum
June
July → travel Europe
August
Picture E11

have a great time
family friends

if Susan

go to Australia

Picture E12

not meet

if not attend

camp last summer
Picture F1
Mina apply for internship UN April

Picture F2
fail English interview
last Friday I meet friends who talk about things with Mina.

Picture F3

friend Sungmin graduate from high school which Mina graduate from

Picture F4
Picture F5

he

friend

who

Mina

problems

with

Picture F6

Hee-Won

friend

who

Mina

talk to

easily
she person

who Mina depend on emotionally problems

friends Mina tell regrets

English interview
if watch CNN often

understand interviewer

questions better

if join

answer

*English Café*

practice speak

better as well
Picture F11

if do → have chance

voice chat practice

English speak English

Picture F12

do better ← if not nervous

interview
if interview successful take

Public Speaking

last semester

Sungmin encourage Mina
Hee-Won prepare apply
internship next year
APPENDIX D

Multiple-Choice Recognition Test: Version A

1. All homework assignments must be turned ___ by next Wednesday.
   (a) in   (b) back   (c) up   (d) out

2. The _______ leaves covered the path and made it quite slippery.
   (a) falling   (b) fallen

3. _____ Mrs. Baker is selling her house and moving to the suburbs.
   (a) a (b)an (c) the (d) Ø

4-5. If he (4)____________ up yesterday morning, he (5) _____________ the first train.
   4. (a) had not hurried                                     5. (a) had not caught
      (b) did not hurry                                             (b) did not catch
      (c) would not have hurried                             (c) could not have caught
      (d) would not hurry                                        (d) could not catch

6. The _______ dogs kept me awake all night.
   (a) barking   (b) barked

7. I hope you can help me _________ what to do about my problem.
   (a) figure out   (b) pick up   (c) make up   (d) look out

8-9. If it (8)_____________ yesterday, we (9)______________ the baseball.
   8. (a) did not rain                                  9. (a) had played
      (b) had not rained                                 (b) played
      (c) would not have rained                     (c) could have played
      (d) would not rain                                 (d) could play

10. Min-ki was born in one of _____ smallest town in the country.
    (a) a (b) an   (c) the   (d) Ø

11. My friends have advised me to find _______ in the phone book.
    (a) attorney   (b) an attorney   (c) the attorney   (d) attorney
12-13. The view was very beautiful. I (12)_____________ some pictures if I (13)____________ a camera.

12. (a) took  
(b) had taken  
(c) would have taken  
(d) would not take

13. (a) had  
(b) had had  
(c) would have had  
(d) would have

14. David’s family thinks that old little Volkswagen has been ___ wonderful car for road trips.

(a) a  
(b) an  
(c) the  
(d) Ø

15. Turn the lights ____. You’re wasting electricity and money.

(a) on  
(b) off  
(c) up  
(d) away

16-17. If he (16)_____________ honest then, I (17)________________ him.

16. (a) was  
(b) had been  
(c) would have been  
(d) would be

17. (a) employed  
(b) had employed  
(c) would have employed  
(d) would employ

18. Mi-so is suffering from ________ heart.

(a) breaking  
(b) broken

19-20. Jane (19)____________ a new job if she (20)___________ in Boston last year.

19. (a) found  
(b) had found  
(c) would have found  
(d) would find

20. (a) stayed  
(b) had stayed  
(c) would have stayed  
(d) would stay
1. Senator John has decided not to run for election and has _______ of the race for health reasons.
   (a) called back   (b) cleared up   (c) dropped out   (d) lay down

2. The _______ waves pushed the surfboard further out to sea.
   (a) breaking   (b) broken

3-4. In-Sung (3) __________ the test he took last week if he (4) __________ very hard last winter.

   3. (a) would have passed   (b) would pass   (c) had passed   (d) passed
   4. (a) would have studied (b) would study (c) had studied (d) studied

5. Herman Melville’s *Moby Dick* is ___ interesting American novel.
   (a) a   (b) an   (c) the   (d) Ø

6. If you make a lot of mistakes, do the homework assignment ______.
   (a) over   (b) up   (c) out   (d) through

7. I was ______ with the movie and I went out.
   (a) boring   (b) bored

8-9. (8) If it __________ then, the game (9) ________________ cancelled.

   8. (a) had not rained   (b) did not rain   (c) would not have rained   (d) would not rain
   9. (a) would not have been (b) would not be   (c) had not been   (d) was not

10. Some doctors think ______ is not safe for everybody.
    (a) a radiation   (b) the radiation   (c) radiation   (d) radiations

11. Please __________ your sunglasses inside the classroom.
    (a) get off   (b) come off   (c) give away   (d) take off
12-13. If I \( (12) \)____________ busy yesterday, I \( (13) \)____________ there.

12. (a) would not have been        13. (a) could go
(b) would not be                  (b) could have gone
(c) had not been                  (c) had gone
(d) was not                      (d) went

14. _____ Statue of Liberty is located in New York Harbor.
(a) a  (b) an  (c) the  (d) Ø

15. ___________ reports are coming in that refugees are being racially abused.
(a) Alarming  (b) Alarmed

16-17. If I \( (16) \)____________ the fact then, I \( (17) \) _____________ it to you.

16. (a) would have known          17. (a) would have told
(b) would know                    (b) would tell
(c) had known                     (c) had told
(d) knew                          (d) told

18. There are extra handouts in the back of the room. Supplies are limited, so please take one before they run ______.
(a) into  (b) over  (c) out  (d) off

19-20. She \( (19) \)________________ the party if she \( (20) \)_______________ a flat tire yesterday.

19. (a) had attended              20. (a) would not have
(b) attended                     (b) had not had
(c) would attend                  (c) did not have
(d) would have attended           (d) would not have had
1. _______________! The light is about to turn red.  
   (a) Eat out   (b) Pay off   (c) Look out   (d) Slip up

2. The _______ waves pushed the surfboard further out to sea.  
   (a) breaking   (b) broken

3-4. We (3) _____________ steak for dinner if we (4) ___________ Jamie was a vegetarian and couldn’t stand meat.  
   (a) would never serve                                    (b) would never have served  
   (c) had never served                                      (d) never served
   (a) would have known                                      (b) would know
   (c) had known                                             (d) knew

5. Herman Melville’s *Moby Dick* is ___ interesting American novel.  
   (a) a   (b) an   (c) the   (d) Ø

6. The teacher will be glad to go ________ the lesson again if you need help.  
   (a) over   (b) up   (c) out   (d) through

7. I was _______ with the movie and I went out.  
   (a) boring   (b) bored

8-9. If you (8)________________ us know you were coming, we (9)_______________ our plans last night.  
   (a) had let                                  (b) would have let
   (b) let                                             (c) would have let
   (c) would have let                         (d) would let
   (d) would let                                  (b) would not change
   (c) had not changed                        (d) did not change

10. Some doctors think _______ is not safe for everybody.  
    (b) a radiation   (b) the radiation   (c) radiation   (d) radiations

11. _________________ this form and then sign your name at the bottom.  
    (a) fill out   (b) hand in   (c) keep up   (d) show up
12-13. My family (12)_____________ to a zoo yesterday if the weather
(13)_____________ nice.
   12. (a) would have gone               13. (a) would be
   (b) would go                                 (b) would have been
   (c) had gone                               (c) had been
   (d) went                                   (d) was

14. _____ Statue of Liberty is located in New York Harbor.
   (b) a   (b) an   (c) the   (d) Ø

15. What time is the mail ___________ during the week?
   (a) delivering   (b) delivered

16-17. What (16)_____________ differently when you were a teenager if you
(17) _______________ then everything that you know now?

   16. (a) would you have done               17. (a) would have known
   (b) would you do                                      (b) would know
   (c) had you done                                    (c) had known
   (d) did you do                                    (d) knew

18. The Drama Club at Seoul High School is putting _____ a play next month.
   (a) in   (b) up   (c) out   (d) on

19-20. Jim’s boss (19)________________ her anymore if she
(20)_____________ him the truth yesterday.

   19. (a) had not trusted               20. (a) would not tell
   (b) did not trust                                (b) had not told
   (c) would not trust                        (c) did not tell
   (d) would not have trusted                  (d) would not have told
BIBLIOGRAPHY


of Pennsylvania, Philadelphia.


Cambridge, MA: MIT Press.


second language acquisition (pp. 413-468). New York: Academic.


Washington, DC: Georgetown University Press.


Martínez-Fernández, A. (2008). Revisiting the Involvement Load Hypothesis:


