CORRECTIVE FEEDBACK, INDIVIDUAL VARIATION IN COGNITIVE CAPACITIES, AND L2 DEVELOPMENT: RECASTS VS. METALINGUISTIC FEEDBACK

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

Jaemyung Goo, M.A.

Washington, DC
April 14, 2011
CORRECTIVE FEEDBACK, INDIVIDUAL VARIATION IN COGNITIVE CAPACITIES, AND L2 DEVELOPMENT: RECASTS VS. METALINGUISTIC FEEDBACK

Jaemyung Goo, M.A.

Thesis Advisor: Alison Mackey, Ph.D.

ABSTRACT

This dissertation explores how the type of structure is related to the effectiveness of different forms of corrective feedback provided during interaction and whether/how individual differences in working memory (WM) and attention control mediate the extent to which L2 learners benefit from corrective feedback. Recasts were compared with metalinguistic feedback in terms of their relative effects on the acquisition of two target structures in an EFL setting: the English that-trace filter and the past unreal conditional. Eighty-three Korean L1 learners of English were randomly assigned to one of three treatment conditions: recasts, metalinguistic feedback, and control. Each learner carried out a series of tasks and tests: a WM span task (operation span), an attention control task (task-set switching), a dyadic intensive treatment activity, oral production tests, and grammaticality judgment tests. Results show that metalinguistic feedback was more effective than recasts at facilitating accuracy in oral production involving the English that-trace filter, whereas recasts were slightly more effective than metalinguistic feedback at promoting the acquisition of knowledge on the past unreal conditional. Also, WM significantly predicted the observed effects of metalinguistic feedback, but not of recasts, on the development of the past unreal conditional, as evidenced in oral production. Attention control, however, was found to be significantly associated with the observed beneficial effects of recasts, but not of metalinguistic feedback, on the learners’ accuracy in oral production of English past unreal conditional sentences. The findings suggest that learners may benefit more from explicit feedback than implicit feedback when learning a
simple rule, but more from implicit feedback than explicit feedback when learning a complex rule. The results also indicate that both WM and attention control may mediate the effectiveness of corrective feedback, but in different ways, that is, their exact roles may differ depending upon the type of feedback.
Acknowledgments

This dissertation is dedicated to my parents, Dal-hoi Goo and Dong-soon Heo, my parents-in-law, Jae-gon Kim and Hong-nim Yang, my brothers, Jaesun and Jinwoo, and, of course, my lovely wife and DeNiro’s mom, Youngjoo Kim, for their infinite support and encouragement throughout my long journey.

My heartfelt thanks go to my advisor and mentor, Alison Mackey, Ph.D., for her continued support and wonderful guidance at each and every step of this process and at every crucial stage of my academic life at Georgetown University. She has been inspirational in various aspects and will always remain so in my future course of life as a researcher and educator. I would also like to deliver my special thanks to Cristina Sanz, Ph.D. and Patrick E. Rebuschat, Ph.D. for insightful feedback and advice on the dissertation during the entire process. Last but not least, as a whole, I am grateful to the Department of Linguistics at Georgetown University for all the opportunities to learn what I have learned and grow as a researcher.

Many thanks,

Jaemyung Goo
Table of Contents

Introduction........................................................................................................................................... 1

Chapter I: Recasts in Interaction Research.......................................................................................... 5

Definitions and Characteristics of Recasts.......................................................................................... 5

Beneficial Effects of, and Controversy Over the Role of, Recasts..................................................... 9

Recent Empirical Studies on Different Types of Feedback................................................................. 13

Methodological Issues in Empirical Studies on Different Types of Feedback.................. 28

Modified Output Opportunities........................................................................................................ 28

Comparisons of Single-Versus-Multiple Forms of Feedback......................................................... 30

Form-Focused Instruction.................................................................................................................. 32

Prior Knowledge................................................................................................................................. 34

Out-of-Experiment Exposure.............................................................................................................. 37

Summary of Problems........................................................................................................................... 38

Factors That Mediate the Beneficial Role of Recasts in L2 Learning............................................. 40

Chapter II: Working Memory Capacity (WMC)............................................................................. 45

Overview........................................................................................................................................... 45

Nature of WM and WMC variation...................................................................................................... 47

WMC and Language Comprehension............................................................................................... 54

WMC and SLA..................................................................................................................................... 56

Goo (In Press)..................................................................................................................................... 66

Research Questions.............................................................................................................................. 67

Method............................................................................................................................................... 67

Results............................................................................................................................................... 69
## Discussion

Limitations and Future Research

WM, Attention Control, and Corrective Feedback

Chapter III: The Present Experiment

Research Questions

Method

Participants

Linguistic Targets

targets

Rationale

Experimental Design

Tasks and Materials

Treatment Task (Rashomon)

Operation Span Task (OSSPAN)

Attention Control Task

Assessment Tasks

Oral Production Test

Grammaticality Judgment Test (GJT)

Procedure

Scoring

OSSPAN and Task-Set Switching

Oral Production/Grammaticality Judgment

Interrater Reliability
## Analysis

Chapter IV: Results

*That*-Trace Filter

Grammaticality Judgment

Summary of GJT Performance on English *That*-Trace Filter

Oral Production

Summary of Oral Performance on English *That*-Trace Filter

Past Unreal Conditional

Grammaticality Judgment

Summary of GJT Performance on Past Unreal Conditional

Oral Production

Summary of Oral Performance on Past Unreal Conditional

Chapter V: Discussion

Research Questions and Overall Results

Discussion on Findings and Related Issues

Summary of Discussion

Chapter VI: Limitations and Future Research

Chapter VII: Conclusion

Notes

Appendix A

Appendix B

Appendix C

Appendix D
Introduction

According to the interactionist perspective (Gass, 1997, 2003; Gass & Mackey, 2006, 2007; Gass, Mackey, & Pica, 1998; Long, 1996, 2007; Mackey, 2007a; Pica, 1994, 1996), interaction during which negotiation for meaning occurs provides favorable grounds for second language (L2) learning to occur as it offers L2 learners potentially beneficial opportunities to receive necessary input or comprehensible input (Krashen, 1982, 1985), produce output through which they can test their hypotheses regarding L2 forms by processing L2 syntactically (Swain, 1985, 1993, 1995, 2005), and attend to the target language code and become aware of the gap between their interlanguage and the target language (Schmidt, 1990, 1993, 1994, 1995, 2001; Schmidt & Frota, 1986), leading them to make necessary efforts to refine and restructure their interlanguage. On the critical role of negotiation for meaning during interaction, Long (1996) argued that “negotiation for meaning, and especially negotiation work that triggers interactional adjustments by the NS or more competent interlocutor, facilitates acquisition because it connects input, internal learner capacities, particularly selective attention, and output in productive ways” (pp. 451-452). Beneficial effects of interaction on the acquisition of L2 lexical and grammatical features are evidenced in previous meta-analyses (e.g., Keck, Iberri-Shea, Tracy-Ventura, & Wambaleka, 2006; Mackey & Goo, 2007; see also Li, 2010; Lyster & Saito, 2010; Russell & Spada, 2006 for meta-analyses on corrective feedback in L2 research). However, as interaction researchers have claimed, the relationship between conversational interaction involving negotiation for meaning and SLA is not a simple causal one, but a complex one that may be affected by various factors (Gass, 1997, 2003; Gass & Mackey, 2006, 2007; Long, 1996; Mackey, 2007a).
Reflecting the complex nature of this relationship, interaction researchers have become particularly interested in specific features of interaction (e.g., noticing, corrective feedback, modified output, negotiation patterns, etc.) that may influence, to varying degrees, the extent to which negotiation for meaning can benefit L2 learning (see for a review and summary of interaction research Gass & Mackey, 2006, 2007; Mackey, 2007a; Mackey & Goo, in press; see also Spada & Lightbown, 2009 for a brief overview of interaction research in L2 classrooms). Especially, an enormous amount of attention has been drawn to recasts, one form of corrective feedback, among other features of interaction, inviting much discussion and generating numerous empirical studies on their nature and relation to L2 learning (Ammar, 2008; Ammar & Spada, 2006; Ayoun, 2001; Braidi, 2002; Carpenter, Jeon, MacGregor, & Mackey, 2006; Doughty & Varela, 1998; Egi, 2007a, 2007b; Ellis, 2007; Ellis, Loewen, & Erlam, 2006; Ellis & Sheen, 2006; Goo, in press; Han, 2002; Ishida, 2004; Iwashita, 2003; Loewen, 2009; Loewen & Nabei, 2007; Loewen & Philp, 2006; Long, 1996, 2007; Long, Inagaki, & Ortega, 1998; Lyster, 1998a, 1998b, 2004; Lyster & Izquierdo, 2009; Lyster & Ranta, 1997; Mackey, 2006; Mackey & Oliver, 2002; Mackey & Philp, 1998; Mackey & Silver, 2005; McDonough, 2007; McDonough & Mackey, 2006; Nassaji, 2009; Nicholas, Lightbown, & Spada, 2001; Oliver, 1995, 1998, 2000; Ohta, 2000; Panova & Lyster, 2002; Philp, 2003; Révész, 2009; Révész & Han, 2006; Sagarra, 2007; Sheen, 2004, 2006, 2007a; Trofimovich, Ammar, & Gatbonton, 2007; Yang & Lyster, 2010). L2 interaction research has successfully shown that recasts facilitate L2 learning (e.g., Ayoun, 2001; Doughty & Varela, 1998; Goo, in press; Han, 2002; Ishida, 2004; Iwashita, 2003; Leeman, 2003; Mackey & Philp, 1998; McDonough & Mackey, 2006; Nassaji, 2009; Ortega & Long, 1997; Sagarra, 2007; Trofimovich et al., 2007) although the role of recasts has been downplayed in L1 acquisition to some extent (e.g., Gordon, 1990; Grimshaw & Pinker, 1989;
Morgan, Bonamo, & Travis, 1995; Pinker, 1989). It should be noted, however, that L1 acquisition researchers have also found positive potential of recasts (e.g., Baker & Nelson, 1984; Bohannon & Stanowicz, 1988; Farrar, 1992; Saxton, 1997; Saxton, Backley, & Gallaway, 2005; Saxton, Kulcsar, Marshall, & Rupra, 1998), that is, controversy still exists concerning the role of negative evidence (e.g., recasts) in L1 acquisition.

More recently, L2 researchers have become increasingly interested in comparing recasts with other forms of feedback such as prompts (e.g., metalinguistic feedback, clarification requests, elicitations, repetitions, etc.) in terms of their relative effects on L2 development in interactional settings (Ammar, 2008; Ammar & Spada, 2006; Ellis, 2007; Ellis et al., 2006; Goo, in press; Loewen & Nabei, 2007; Lyster, 2004; Lyster & Izquierdo, 2009; McDonough, 2007; Nassaji, 2009; Sheen, 2007a; Yang & Lyster, 2010). Other empirical studies on corrective feedback in non-interactional settings or in different modes have also been conducted and contributed to our understanding of how different types of feedback affect L2 learning (e.g., Carroll & Swain, 1993; Sheen, 2007b, 2010; Storch & Wigglesworth, 2010; see also Li, 2010; Lyster & Saito, 2010; Russell & Spada, 2006 for meta-analyses of findings of L2 feedback research including studies conducted in non-interactional settings as well as those in interactional settings).

However, one potentially crucial factor that may determine the extent to which corrective feedback (e.g., recasts, metalinguistic feedback, etc.) facilitates L2 learning, individual differences in working memory capacity (WMC), has not been fully considered in this line of research on the relative efficacy of one type of feedback over another (see Goo, in press for an exception). WM as a critical cognitive construct has been researched extensively and thoroughly in the field of cognitive psychology (see Baddeley, 2007; Conway, Jarrold, Kane, Miyake, &
Towse, 2007a; Cowan, 2005; Jarrold & Towse, 2006; Miyake, 2001; Miyake & Shah, 1999 for various theories and models of WM and its theoretical developments) and found to be significantly correlated with higher-order cognitive behaviors (e.g., reading, reasoning, general fluid intelligence\(^1\), etc.). Similarly, WM has been much discussed and studied in the field of SLA with respect to its relation to L2 learning and relevant research (e.g., Geva & Ryan, 1993; Goo, 2010, in press; Harrington & Sawyer, 1992; Havik, Roberts, van Hout, Schreuder, & Haerker, 2009; Juffs, 2004, 2005; Kormos & Sáfár, 2008; Mackey, Adams, Stafford, & Winke, 2010; Mackey, Philp, Egi, Fujii, & Tatsumi, 2002; Mackey & Sachs, in press; Miyake & Friedman, 1998; Rai, Loschky, Harris, Peck, & Cook, 2011; Robinson, 2005b; Sagarra, 2007, 2008; Sagarra & Herschensohn, 2010; Trude & Tokowicz, 2011; Tokowicz, Michael, & Kroll, 2004; Trofimovich, Ammar, & Gatbonton, 2007; Walter, 2004; see also Juffs & Harrington, 2011; Williams, in press for reviews of research on WM and L2 learning). However, only a few studies have attempted to investigate whether and how WMC affects the efficacy of corrective feedback (e.g., recasts) on the actual development of L2 linguistic forms (Goo, in press; Mackey et al., 2002; Mackey & Sachs, in press; Sagarra, 2007; Trofimovich et al., 2007). In fact, Goo’s (in press) study is the only empirical attempt to explore the relationship between WMC and the type of corrective feedback in interaction-driven L2 learning. This dissertation is designed as an expanded version of Goo’s study with one additional target structure and one additional cognitive variable assumed to be associated with WMC. Discussed in the following sections are previous research on recasts and relevant methodological issues, followed by WM research and the present dissertation study.
Chapter I: Recasts in Interaction Research

Definitions and Characteristics of Recasts

Based on L1 literature on recasts, Long (1996) defined recasts as “utterances that rephrase a child’s utterance by changing one or more sentence components (subject, verb, or object) while still referring to its central meanings” (p. 434). Emphasizing more implicit and incidental characteristics of recasts, Long (2007) later defined a corrective recast as “a reformulation of all or part of a learner’s immediately preceding utterance in which one or more nontargetlike (lexical, grammatical, etc.) items is/are replaced by the corresponding target language form(s), and where, throughout the exchange, the focus of the interlocutors is on meaning, not language as object” (p. 77). In L2 research, recasts have been generally defined as reformulated utterances with one or more changes made to L2 learners’ nontargetlike utterances while maintaining their original meaning (see also Ellis & Sheen, 2006; Nicholas et al., 2001 for more definitions). In other words, recasts are more targetlike versions of learners’ nontargetlike utterances. They can be provided, and have been operationalized in L2 research, in various forms and patterns (e.g., recasts with or without phonological enhancement, partial or full recasts, etc.) with potentially differing degrees of impact on L2 development (see Loewen & Philp, 2006 for more information).

Regardless of somewhat different renditions of recasts in operationalization, two crucial contributing factors to beneficial effects of recasts on L2 learning are semantic contingency and an immediate juxtaposition of learners’ erroneous output and corrective recasts (Long, 1996, 2007). L2 learners likely understand all or part of their interlocutor’s message because it is a reformulated version of the learner’s nontargetlike utterance. This promotes semantic transparency, which in turn enables learners to utilize more attentional resources to focus on
form and form-function mapping. Long (2007) claimed, “Recasts convey needed information about the target language in context [italics in original], when interlocutors share a joint attentional focus [italics in original], and when the learner already has prior comprehension [italics in original] of at least part of the message, thereby facilitating form-function mapping” (p. 77). He further noted that “learners are vested [italics in original] in the exchange, as it is their message that is at stake, and so will probably be motivated [italics in original] and attending [italics in original], conditions likely to facilitate noticing [italics in original] of any new linguistic information in the input” (pp. 77-78). Doughty and Varela (1998) also suggested that recasts are “potentially effective since the aim is to add [italics in original] attention to form to a primarily communicative task rather than to depart [italics in original] from an already communicative goal in order to discuss a linguistic feature” (p. 114). Also, an immediate juxtaposition of the learner’s erroneous output and corrective recasts, by drawing “learners’ attention to mismatches between input and output” (Long & Robinson, 1998, p. 23), renders highly plausible a cognitive comparison of the two forms that are in contrast, which leads the learner to perceive the form contained in a recast as a correct alternative to their erroneous form. In his direct contrast hypothesis, Saxton (1997) argued that “negative evidence is especially well adapted for highlighting not only the existence of such contrasts, but moreover, for revealing which of two linguistic forms should be retained and which rejected” (p. 155). Similarly, Doughty (2001) suggested that immediately contingent recasts are promising elements that may lead to effective cognitive comparisons, which in turn makes form-meaning-function mapping highly likely to occur during speech processing.

Although recasts are generally considered as one type of negative evidence/feedback (see Long & Robinson, 1998 for a relevant figure, according to which recasts are classified as
implicit negative evidence), the exact nature of recasts has been debated as a somewhat complicated issue because recasts are targetlike reformulated utterances provided in the form of positive evidence or models. Ellis and Sheen (2006) suggested that recasts consist of either positive evidence only or a combination of negative and positive evidence, depending on whether learners notice its corrective nature. That is, if recasts are interpreted as corrective, then they constitute negative evidence as well as positive evidence. If learners are not aware of its corrective nature, however, recasts are interpreted as providing positive evidence only. However, their characterization of recasts based entirely on learner perception and interpretation does not seem to reflect the general mechanism of recasts. Recasts are provided in response to learners’ non-targetlike utterances and are intended to impart their non-targetlikeness to the learners, which is what makes recasts negative evidence regardless of learner perception. In other words, negative evidence is an intrinsic part of what is provided through recasts, and how recasts are perceived is a somewhat different matter. Regarding learner perception of recasts, data from verbal protocols (immediate reports and stimulated recall protocols) in Egi’s (2007b) study indicated that when recasts were attended to (or noticed), they were interpreted as negative evidence (or a combination of negative and positive evidence) more often than as positive evidence regardless of the type of linguistic target (e.g., morphosyntax and lexis). However, based on the results of her study designed to investigate the nature of recasts, Leeman (2003) interpreted her results as suggesting that beneficial effects of recasts on L2 learning are due mainly to enhanced positive evidence, and that negative evidence may not be a critical factor. Considering Leeman’s study is the only attempt to examine the extent to which each component (positive and negative evidence) of a recast contributes to L2 learning, the precise nature of a
recast in this respect definitely merits further research although, of course, it requires quite a sophisticated experimental manipulation to separate positive/negative evidence from recasts.

With regard to the issue of implicitness, Long (2007) claimed that “a recast is a discourse move that is by definition implicit [italics in original]…designed not to interrupt speakers’ focus on message…” (p. 99). Ellis and Sheen (2006), however, suggested that recasts can be viewed as more or less explicit or implicit, but not entirely implicit, depending upon linguistic features utilized to encode recasts (e.g., intonation, length, number of changes, etc.) and the context in which they are used (e.g., didactic vs. communicative, ESL/FL classrooms vs. immersion), both of which may affect the noticeability of, and learners’ interpretations of, recasts. Long’s view of “implicit” reflects the nonintrusive nature of recasts in meaning-based interaction (a discourse-level definition), whereas Ellis and Sheen’s (2006) interpretation of explicitness and implicitness is based on how and in which contexts recasts are actually used and noticed (a usage-based micro-level definition). That is, their definitions are based on different functional aspects of recasts. As regards how recasts can be viewed during interaction, Egi (2007b), mentioned earlier, found that recasts were interpreted as responses to content when they were long and when they involved three or more changes (that is, somewhat implicit recasts), whereas shorter recasts along with one or two corrective changes (that is, more explicit recasts) were more likely to be perceived as linguistic evidence (positive, negative, or both), indicating that the length and number of changes in recasts may influence L2 learners’ interpretations of recasts. Similarly, Philp (2003) in her laboratory study showed that shorter recasts (five or fewer morphemes in her study) were more likely to be noticed by L2 learners than longer recasts. Also, in her analysis of data from ESL/FL classroom learning contexts, Sheen (2006) found that word- or short phrase-length recasts led to more uptake than clause-level recasts. Loewen and Philp (2006), drawing on
adult ESL classroom data (17 hours of meaning-based interaction), described characteristics of recasts, in detail, as emphasis, focus, segmentation, intonation, prosodic emphasis, and degree of explicitness, in relation to their relative contributions to the efficacy of recasts on L2 learning, and showed that intonation, morpheme length, and number of changes were the most relevant factors that mediated the effects of recasts on learner performance.

**Beneficial Effects of, and Controversy Over the Role of, Recasts**

L2 research to date has empirically shown that recasts promote L2 learning (e.g., Ayoun, 2001; Doughty & Varela, 1998; Han, 2002; Ishida, 2004; Iwashita, 2003; Leeman, 2003; Long et al., 1998; Mackey & Philp, 1998; McDonough & Mackey, 2006; Nassaji, 2009; Sagarra, 2007; Trofimovich et al., 2007; see also Long, 2007 for a review of research on recasts and Mackey & Goo, 2007 for a meta-analysis where they found large mean effect sizes on all three posttests). For instance, Mackey and Philp (1998), one of the earliest L2 studies on the role of recasts in L2 development, found the positive effects of recasts on the acquisition of English question formation (for similar findings, see McDonough & Mackey 2006 on Thai EFL learners’ English question development). Doughty and Varela (1998) also showed recasts were facilitative of the development of English simple past and conditional; they used a more explicit version of recast, corrective recasting (or focused recasting), a recast preceded by a repetition. Han (2002) examined the effects of recasts on tense consistency in L2 English output and found that recasts had a positive impact on her learners’ ability to maintain tense consistency in their oral and written output. Moreover, learner awareness of tense consistency in the recast group increased to a great degree, compared to the non-recast group. Also, Ishida (2004) in her study, employing a time-series design, observed a significant accuracy increase in her Japanese as a foreign language learners’ use of Japanese aspectual form –te i-(ru) after the intensive recasting began. Long et al.
(1998, also in Ortega & Long, 1997) found the beneficial role of recasts in the acquisition of Spanish adverb placement (see their discussion for more information on the results). The positive effects of recasts on the acquisition of Spanish noun-adjective gender-number agreement were evidenced in Leeman’s (2003) study. Recasts delivered via computer have also been found to be beneficial to L2 learning (e.g., Ayoun, 2001; Sagarra, 2007; Trofimovich et al., 2007). More recently, Nassaji (2009) showed that recasts led to more successful post-interaction corrections than did elicitation moves.

In addition, Mackey and Goo (2007) provided convincing evidence for the beneficial role of recasts in L2 learning. They meta-analyzed findings of interaction studies published between 1990 and 2006 (June). They found large mean effect sizes for recasts on all three posttests ($M = .96, SD = 1.04$ for immediate posttests, $M = 1.69, SD = 1.13$ for short-term delayed posttests, and $M = 1.22, SD = .85$ for longer-term delayed posttests). Especially, the mean effect size for recasts on the immediate posttests was significantly higher than those for negotiation moves (e.g., clarification requests, confirmation checks, comprehension checks) and metalinguistic feedback. As Mackey and Goo (2007) cautioned, however, more empirical studies on recasts seem necessary to gain a clearer picture of the impact of recasts (and possibly their relative effectiveness over other types of corrective feedback).

As briefly described here, empirical studies on recasts to date (and a meta-analysis) have successfully demonstrated that recasts play a facilitative role in L2 learning. However, some classroom-based descriptive research has discussed recasts in terms of the extent to which recasts can lead to modified output or uptake, casting doubt on the effectiveness of recasts in L2 learning due to relatively fewer instances of learner uptake following recasts (Lyster, 1998a, 1998b; Lyster & Ranta, 1997; Panova & Lyster, 2002). Lyster and Ranta (1997) conducted a
study designed to investigate the relationship between corrective feedback and learner uptake. They defined uptake as “a student’s utterance that immediately follows the teacher’s feedback and that constitutes a reaction in some way to the teacher’s intention to draw attention to some aspect of the student’s initial utterance” (p. 49). It was observed that even though teachers frequently used recasts to provide corrective feedback, recasts did not lead to as much uptake as did other forms of feedback such as elicitation and metalinguistic feedback, which was reiterated in Lyster’s (1998a, 1998b) subsequent papers (see also Panova & Lyster, 2002 for similar results). However, unlike what Lyster and his colleagues have maintained, research based on classroom data has shown that recasts do in fact lead to substantial uptake in some instructional settings where teachers and learners focus on language as form (Sheen, 2004) and interactional contexts where explicit language-focused exchanges are involved (Oliver & Mackey, 2003). Oliver (1995) also showed that child L2 learners incorporated more than 33% of the recasts that they received when given the opportunity and when it was appropriate to do so (see also Braidi, 2002 for a similar result). Additionally, in her longitudinal study of Japanese as a foreign language learners Ohta (2000) found that learners produced private speech in response to recasts directed at the whole class or other students, implying that L2 learners do in fact incorporate recasts above and beyond the surface level.

Furthermore, Lyster’s overall skepticism on the utility of recasts has engendered much criticism from quite a few SLA researchers (e.g., Gass, 2003; Long, 2007; Mackey & Philp, 1998). Criticizing Lyster and Ranta’s arguments, Mackey and Philp (1998) pointed out uptake (in Lyster and Ranta’s sense) may not be an appropriate measure of effectiveness of recasts, and that delayed effects should be properly and carefully measured if one is to find evidence of effectiveness of recasts in terms of development or acquisition. They suggested that learners’
immediate responses may be *red herrings*. In their study on the relationship between responses to recasts and ESL question development, McDonough and Mackey (2006) found that primed production operationalized as “a learner’s use of the question form provided in the recast to ask a new question” (p. 705) was a statistically significant predictor of learner performance on English question formation, but immediate repetition (one type of uptake in Lyster and Ranta’s sense) was not. In his criticism of Lyster’s skepticism, Long (2007) argued that Lyster’s claims and concerns regarding the utility of recasts are simply unwarranted. He made strong counterarguments, point by point, to Lyster’s skepticism of recasts, by providing supportive evidence (logical as well as empirical evidence) for the positive value and potential utility of recasts. Long discussed several possible reasons for ambiguity of recasts (including those related to instructional contexts), the questionable validity of immediate uptake as an indication of acquisition in relation to the issue of deployment vs. acquisition, the nature of recasts, the coarse-grained definition of uptake used in Lyster’s and Lyster and his colleagues’ studies, some questionable interpretations of the classroom data made in their studies, and the benefits of the unobtrusiveness of recasts as opposed to the inescapable obtrusiveness of explicit feedback.

Nevertheless, triggered predominantly by Lyster’s skepticism over the role of recasts in L2 development, L2 interaction researchers have shown their continued interests in comparing recasts with prompts such as metalinguistic feedback, clarification requests, elicitations, and repetitions (Ammar, 2008; Ammar & Spada, 2006; Ellis, 2007; Ellis et al., 2006; Goo, in press; Loewen & Nabei, 2007; Lyster, 2004; Lyster & Izquierdo, 2009; McDonough, 2007; Nassaji, 2009; Sheen, 2007a; Yang & Lyster, 2010). In fact, the issue of the relative efficacy of one form of feedback over another on L2 learning has emerged as a potentially promising area of inquiry in interaction research. For example, clarification requests (e.g. *I don’t understand, what?*, *Huh?*, *I'm not sure, what?*,
etc.) have been examined not only in terms of their own impact on L2 learning (e.g., McDonough, 2005; Nobuyoshi & Ellis, 1993; Takashima & Ellis, 1999), but also in comparison with other types of feedback including recasts (e.g., Loewen & Nabei, 2007; Lyster & Izquierdo, 2009; McDonough, 2007). Another type of corrective feedback that has often been compared with recasts is metalinguistic feedback. Metalinguistic feedback is defined as “either comments, information, or questions related to the well-formedness of the student’s utterance, without explicitly providing the correct form” (Lyster & Ranta, 1997, p. 47). However, it has been operationalized in various ways. For instance, Ellis et al. (2006) operationalized it as a repetition followed by metalinguistic information (also on the development of English past tense in Ellis’s (2007) study). Also, in Sheen’s (2007a) study it was operationalized as recasts followed by metalinguistic information. She termed it metalinguistic correction. This type of feedback either alone or in combination with other forms of feedback that provide L2 learners with opportunities to repair their erroneous utterances has also been compared with recasts (e.g., Ammar, 2008; Ammar & Spada, 2006; Ellis, 2007; Ellis et al., 2006; Goo, in press; Loewen & Nabei, 2007; Lyster, 2004; Sheen, 2007a; Yang & Lyster, 2010). In the next section, recent empirical studies on the relative efficacy of different types of corrective feedback on L2 development are briefly summarized, followed by an in-depth discussion of their methodological issues in the subsequent section.

**Recent Empirical Studies on Different Types of Feedback**

Lyster (2004) investigated the effectiveness of prompts and recasts along with form-focused instruction (FFI). Four groups were compared in terms of the development of French gender assignment: FFI with recasts, FFI with prompts, FFI only, and comparison. The FFI-prompt group performed significantly better than the FFI-recast group in both posttests of the
written tasks (binary-choice and text completion). But the performance of the FFI-prompt group on both posttests of the oral tasks (object identification and picture description) did not show any significant differences as compared to that of the FFI-recast or FFI-only groups although the FFI-prompt group showed significantly better performance than the comparison group in both posttests; however, the FFI-prompt was the only group that performed significantly better than the comparison group in the first posttest. Lyster argued, based on the results, that FFI was more effective with prompts than with recasts or no corrective feedback. However, it can also be argued that his study provided no direct evidence, but rather only indirect evidence for the superiority of prompts as opposed to recasts because form-focused instruction, regardless of the type of feedback, likely played a moderating role. Also, the difference between the FFI-prompt and the FFI-recast groups did not reach any significance level in either of the oral posttests although the FFI-prompt is the only group that performed significantly better than the comparison group in the first posttest. Thus, which type of feedback is more effective is still an open empirical question; this also clearly shows that the choice of a dependent measure is another variable that we should consider as it may affect the results of interaction research (e.g., Mackey & Goo, 2007; see also Norris & Ortega, 2000, 2003 for a similar line of discussion).

Similarly, Ammar and Spada (2006) compared prompts (e.g., elicitation, metalinguistic feedback, and repetition) and recasts in terms of their effects on the development of English third-person singular possessive determiners (*his* and *her*). They also examined the role of proficiency in the efficacy of recasts and prompts. All sixty-four students participated in one instruction session (same for all three groups: recasts, prompts, control) and eleven practice sessions over a period of 4 weeks, during which they received the actual instructional treatment depending on their group affiliation (recasts, prompts, no feedback). Results, based on two
dependent variable measures (a passage correction and an oral picture description task), showed that overall, the prompt group performed better than the recast group (the differences were significant between the two experimental groups except the immediate posttest in the oral picture description task). They also reported that low-proficiency learners (those with less than 50% in accuracy in the pretests) had benefited more from prompts than from recasts, whereas high-proficiency learners (those with more than 50% in accuracy in the pretests) had benefited equally from both types of feedback. They suggested that explicitness and uptake opportunities were attributable to the superiority of prompts. Further support for prompts over recasts was reported in Ammar’s (2008) study in which data from Ammar and Spada’s study (oral picture description data) were reanalyzed in terms of developmental stages; prompts are more effective in helping learners move up to advanced developmental stages than recasts. In addition, Ammar reported the results of reaction time in a computerized fill-in-the-blank test, which showed that the prompt group was significantly faster than the recast group in performing the posttest.

McDonough (2007) reported the findings of an empirical study that had compared the developmental outcomes associated with two types of interactional feedback, recasts and clarification requests. Whereas previous studies compared recasts with more explicit forms of interactional feedback, the focus of comparison in her study was between recasts and another implicit type of feedback. The study was designed to investigate a) whether recasts are associated with development despite infrequent learner responses, and b) whether implicit feedback that does elicit responses (i.e. clarification requests) is more effective than recasts. McDonough’s study adopted this predictive framework to compare the impact of recasts with that of clarification requests on the emergence of simple past activity verbs, which are presumed to emerge later than achievement or accomplishment verbs. Two-way information exchange and
information gap tasks were used in the treatment sessions and one-way information gap tasks in the testing sessions. Learner development was measured on the basis of the number of past activity verbs produced at the posttests. The study found that both clarification requests and recasts facilitated the emergence of past tense activity verbs. However, no significant difference was found between the recast and clarification request groups.

Sheen’s study (2007a) addressed differential effects of two feedback types on the acquisition of English articles and the extent to which individual differences in language aptitude and attitude towards error correction influence the effectiveness of corrective feedback. Recasts and metalinguistic correction were compared in her study. As mentioned earlier, metalinguistic correction was operationalized as feedback involving recasts plus metalinguistic explanations. Two narrative tasks (story retelling) were carried out over two treatment sessions. For each testing session (pretest, posttest and delayed posttest), three tests were administered: a speeded dictation test, a writing test, and an error correction test. In addition, an aptitude test was administered prior to the pretest session. Results showed that the metalinguistic group outperformed both the recast group and the control group, whereas there was no significant difference between the recast and control groups. In the metalinguistic group, there was a significantly positive relationship between students’ immediate and delayed gain scores on the one hand and their language analytical ability on the other. The immediate gains in the metalinguistic group were also positively correlated with learners’ attitude towards errors and error correction. In contrast, the gain scores in the recast group were not significantly related to learners’ ability for language analysis or their reported feelings about error correction. It should be noted, however, that her metalinguistic group received recasts as well as metalinguistic information, that is, this form of metalinguistic treatment was much more explicit than
metalinguistic information alone. Thus, her results should be interpreted with caution, which is further discussed in a later section.

Another study by Loewen and Nabei (2007) investigated the effects of corrective feedback on the development of English question formation during meaning-focused tasks. They examined the differential effects of three types of feedback: metalinguistic feedback, recasts, and elicitations (i.e. clarification requests and requests for a repetition of an erroneous question) on learner performance. Sixty-six Japanese EFL learners attending a Japanese university participated in the study. Five groups were formed: metalinguistic feedback, recast, elicitation, no feedback, and control. A spot-the-difference and a guess-the-storyline task were utilized during the treatment. Two types of dependent variable measures were administered: grammaticality judgment tests (GJTs) and oral production tests. Their results showed that there was no significant difference among the groups in the untimed GJT (based on an ANCOVA), but found a between-group difference to be significant in the timed GJT (based on an ANOVA on gain scores). The feedback groups performed significantly better than the no feedback group and the control group, but no significant difference was observed among the feedback groups. In addition, no significant effect of feedback on English question formation was evidenced in the oral production task, which is quite surprising considering earlier research findings that suggest interactional feedback contributes to English question development in oral production (e.g., Mackey & Philp, 1998; Mackey & Oliver, 2002; Mackey & Silver, 2005; McDonough, 2005; McDonough & Mackey, 2006).

Ellis et al. (2006) compared two types of corrective feedback (recasts and metalinguistic feedback) in terms of their effectiveness on the acquisition of English past tense morpheme –ed. Learner performance was measured via an oral imitation test, an untimed GJT, and a
metalinguistic knowledge test. The results of the oral imitation and the untimed GJT showed that the metalinguistic group performed significantly better than the recast group (and the control group) in the delayed posttest, although no significant between-group difference was found in the immediate posttest in either test measure. Also, the metalinguistic group was significantly better at generalizing the past tense morpheme to new items than the recast group, displaying significantly better performance in both the oral imitation task and the untimed GJT, but only in the delayed posttests. Interesting but not discussed in their article is the observation that the control group performed significantly better than did the recast group on the grammatical items in the untimed GJT (delayed posttest). A likely explanation is that out-of-experiment exposure to the target structure occurred between the immediate posttest and the delayed posttest (e.g., studying, receiving input related to the structure outside the experimental settings) affecting the control group’s performance, thus calling into question the validity of the results regarding the significant difference between the control group and the recast group. This issue is also discussed in a later section.

Particularly interesting in relation to the current study is the one conducted by Ellis (2007) which examined differential effects of two types of corrective feedback on the acquisition of two different structures, the past tense morpheme –ed and the comparative –er. Three types of dependent variable measures were employed as in Ellis et al.’s (2006) study: oral imitation test, untimed GJT, and metalinguistic knowledge test. Results showed that recasts did not lead to differential effects on the acquisition of the two structures, whereas metalinguistic feedback had differential effects on the acquisition of the two target structures. This was evidenced, in particular, in the results of the oral imitation task with ungrammatical sentences and the untimed GJT with ungrammatical sentences. Metalinguistic feedback had a greater immediate impact on
the acquisition of the comparative (i.e., performance on the immediate posttest), but a greater delayed impact on the acquisition of the past tense morpheme –ed (i.e., performance on the delayed posttest) in the oral imitation task with ungrammatical sentences. Also, metalinguistic feedback was more effective for the comparative than for the past tense morpheme in the untimed GJT with ungrammatical sentences. Despite these interesting findings, it should be noted, as Ellis also mentioned, prior knowledge was not controlled, and hence affected the overall picture of his findings (see the next section for more details on this and other methodological issues).

Lyster and Izquierdo (2009) examined the relative effects of recasts and prompts on the development of French grammatical gender. Twenty-five university students participated in a form-focused instructional treatment for approximately 3 hours over 2 weeks. Then, they were randomly assigned to each of two treatment groups: recasts (recasts only without modified output opportunities) and prompts (clarification requests with modified output opportunities, sometimes followed by repetition with rising intonation when a clarification did not elicit a correct repair). The participants carried out three interactional activities (object-identification task, picture-description task, and riddles) in each of two treatment sessions and two oral production tasks (object-identification and picture-description tasks) and a computerized binary-choice test designed to measure accuracy and reaction time in each of three test sessions (pretest, posttest, and delayed posttest). Results showed a significant main effect for Time (pretest-posttest-delayed posttests), but a nonsignificant main effect for Group (Recast vs. Prompts), and a nonsignificant interaction effect between Time and Group in all three dependent variable measures and the reaction-time measure. The researchers attributed the nonsignificant between-groups findings to the heightened level of explicitness of both feedback types due to consistent
and intensive provision of feedback in their lab-based study setting which also entailed individualized attention, form-focused instructional treatment, and form-focused orientation of their learners.

Nassaji (2009) also compared recasts with elicitations (prompts in Lyster’s sense) in terms of their effects on L2 learning through individualized learner-specific tests. Forty-two adult ESL learners participated in a dyadic picture-sequencing task with one of their two NS teachers. Each learner wrote a written description of the story represented by the pictures, then were asked to orally describe the story that they wrote based on the pictures during the interaction with the NS teacher who provided corrective feedback on the learner’s ill-formed utterances. Immediately after the interaction, the learners were instructed to identify and correct ill-formed utterances in their original written description (posttest). They performed the same identification and correction activity 2 weeks later (delayed posttest). It was found that learners were more likely to successfully identify and correct their ungrammatical utterances after receiving recasts (40.7%) than elicitations (28.1%). Also evidenced was that more explicit forms of feedback, in both types of feedback (recasts vs. elicitations), led to more successful corrections than implicit forms of feedback. However, the results were based on descriptive statistics, and thus more research in this framework seems necessary in order to provide supportive evidence for his findings.

Yang and Lyster (2010) conducted a quasi-experimental study in which a total of 72 Chinese EFL learners carried out two dictogloss tasks, one question-and-answer task, and one picture-cued narrative task over a 2-week period. Recasts were compared with prompts including clarification requests, metalinguistic clues, repetitions, and elicitations in terms of their effects on the acquisition of English regular and irregular past tense. The researchers found that there was
no significant difference between the prompt and recast groups. However, there was a significant difference between the prompt and control groups in the oral production of regular past tense. In fact, it was the only significant between-subjects difference, and no significant difference was found among the groups in any other measures. Despite the lack of direct evidence, the researchers interpreted the results, based on their within-subjects analyses that appeared to be in favor of the prompt condition compared to the recast condition, as suggesting that prompts are more effective than recasts.

More recently, Goo (in press) also implemented a quasi-experimental study designed to compare the effects of recasts and metalinguistic feedback on the acquisition of the English *that*-trace filter. Fifty-four Korean EFL learners participated in two one-way information gap activities in a classroom setting and received either recasts or metalinguistic feedback depending upon their group affiliation. A control group carried out the pretest and posttest only, but no interactional activity. Two types of dependent variable measures were developed and administered to the learners, a timed grammaticality judgment test and a written production test. Results showed that both experimental groups significantly outperformed the control group, and that there was no significant difference between the two experimental groups. These results were found on learner performance on both dependent measures. Unlike most studies in this line of research, modified output opportunities were not allowed in Goo’s study, which he suggested may have affected the results. Goo’s study also examined whether and how WMC mediate the effectiveness of corrective feedback. This portion of the study is elaborated on in a later section on WMC and SLA.

Overall, empirical studies on the relative efficacy of one feedback type over another on L2 learning have reported somewhat mixed findings (see Table 1 for a brief summary). Seven
studies were interpreted in their own reports as providing evidence that recasts may not be as effective as other feedback types consisting of two or more forms of feedback in most cases (Ammar, 2008; Ammar & Spada, 2006; Ellis, 2007; Ellis et al., 2006, Lyster, 2004; Sheen, 2007a; Yang & Lyster, 2010). Four studies showed nonsignificant differences between recasts and other feedback types and were interpreted as such (Goo, in press; Loewen & Nabei, 2007; Lyster & Izquierdo, 2009; McDonough, 2007). However, Nassaji (2009) reported that recasts had led to a higher rate of corrections on the immediate and delayed posttests than elicitation moves (prompts in Lyster’s sense).
### Table 1

**Recent Interaction Studies Comparing Different Types of Feedback**

<table>
<thead>
<tr>
<th>Study</th>
<th>Participants</th>
<th>Target structure(s)</th>
<th>Design</th>
<th>Dependent variable measures</th>
<th>Reported results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lyster (2004)</td>
<td>179 5th graders in a French immersion school</td>
<td>French grammatical gender</td>
<td>1. FFI + recasts&lt;br&gt;2. FFI + prompts&lt;br&gt;3. FFI only&lt;br&gt;4. Control</td>
<td>Binary-choice test, text-completion test, object-identification test, picture-description test</td>
<td>1. The FFI-prompt group performed significantly better than the FFI-recast group in the written posttests (binary-choice and text-completion).&lt;br&gt;2. Evidence in favor of FFI-prompts in the oral tests</td>
</tr>
<tr>
<td>Ellis, Loewen, &amp; Erlam (2006)</td>
<td>34 ESL students in a private language school in New Zealand</td>
<td>English Past tense morpheme -ed</td>
<td>1. Recast group&lt;br&gt;2. Metalinguistic explanation&lt;br&gt;3. Control</td>
<td>Oral imitation test, untimed grammaticality judgment test, metalinguistic test</td>
<td>1. No between-group difference in the immediate posttests&lt;br&gt;2. In the delayed posttests, the metalinguistic group performed significantly better than the recast and control groups in the oral imitation and the untimed grammaticality judgment tests.&lt;br&gt;3. The metalinguistic group was significantly better at generalizing the past tense morpheme to new items than the recast group (only evidenced in the delayed posttests).</td>
</tr>
<tr>
<td>Ammar &amp; Spada (2006)</td>
<td>64 grade 6 intensive ESL learners in the Montreal area</td>
<td>3rd person possessive determiners (his and her)</td>
<td>1. Recast group&lt;br&gt;2. Prompt group&lt;br&gt;3. Control</td>
<td>Passage correction, Picture description</td>
<td>1. Overall, the prompt group performed better than the recast group (the differences were significant between the two experimental groups except the immediate posttest in the oral picture</td>
</tr>
<tr>
<td>Study</td>
<td>Participants</td>
<td>Treatment</td>
<td>Oral production</td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------------------------</td>
<td>---------------------------------------</td>
<td>---------------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>McDonough (2007)</td>
<td>74 Thai EFL learners (first-year university students)</td>
<td>Simple past activity verbs</td>
<td>1. Recast group</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2. Clarification request group</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3. Interaction with no feedback</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Treatment: Two-way information exchanges and information gap tasks</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1. Both clarification request and recast groups facilitated the emergence of past tense activity verbs.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2. No advantage for clarification requests over recasts.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sheen (2007a)</td>
<td>80 adult ESL learners in the American Language Program of a community college</td>
<td>English articles</td>
<td>1. Recast group</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2. Metalinguistic correction (recast followed by metalinguistic information)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3. Control</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Treatment: Two narrative tasks in two sessions (30-40 minutes for each task)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1. The metalinguistic group outperformed both the recast and control groups</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2. For the metalinguistic group, significant positive relationship between the immediate and delayed gain scores on the one hand and their language analytical ability on the other.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3. The immediate gains in the metalinguistic group were also positively correlated with learners’ attitude towards error/error correction.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>4. For the recast group, no significant relationship between the gain scores and the learners’ ability for language analyses, nor their reported feelings about error correction.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2. Recast</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3. Elicitation</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Grammaticity Judgment test (timed and untimed), Oral</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1. No significant between-group difference in the untimed grammaticity judgment test and the oral production test</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study</td>
<td>Participants</td>
<td>Treatment</td>
<td>Pre-/Post-test Measures</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------------------</td>
<td>-------------------------------------------------------------------------------</td>
<td>---------------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ellis (2007)</td>
<td>34 adult ESL learners in a private language institute in New Zealand</td>
<td>Two half-hour storytelling activities over two days for the past tense -ed and description activities for the comparative</td>
<td>Oral imitation test, Untimed grammaticality judgment test, Metalinguistic knowledge test</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ammar (2008)</td>
<td>64 grade 6 intensive ESL learners in the Montreal area</td>
<td>One instructional session and 11 practice sessions (reported only as communicative activities) over 4 weeks. Each session took 30-45 minutes.</td>
<td>Oral picture description (developmental stages) Computerized fill-in-the-blank (reaction time)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lyster &amp; Izquierdo</td>
<td>25 university students learning French grammatical</td>
<td>One spot-the-difference task and one guess-the-storyline task (30 minutes for the entire treatment session)</td>
<td>Object identification task, Picture description</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- **4. Interaction with no feedback**
- **5. Control**

- **2. Significant difference between the feedback groups and no feedback group (and control group) in the timed one, but no significant difference among the feedback groups**

- **1. Recasts did not lead to differential effects on the acquisition of the two structures.**
- **2. Metalinguistic feedback had a differential effect on the acquisition of the two target structures.**

- **1. All learners in the prompt group moved up to the post-emergence (stages 5-7), whereas 65% of the recast group learners did.**
- **2. All prompt group learners were able to stay at the post-emergence stage, but one third of the learners in the recast and control groups failed to maintain their stage increase achieved on the immediate posttest.**
- **3. The prompt group was significantly faster than the recast group in performing a computerized fill-in-the-blank task.**

- **1. Significant pre-to-post developments/changes in all measures.**
<table>
<thead>
<tr>
<th>Study</th>
<th>Participants</th>
<th>Language Focus</th>
<th>Targets</th>
<th>Activity 1</th>
<th>Activity 2</th>
<th>Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nassaji (2009)</td>
<td>42 adult ESL learners</td>
<td>French gender</td>
<td>Multiple targets</td>
<td>1. Recast group 2. Elicitation group 1 picture-sequencing task for 10 to 15 minutes</td>
<td>Identifications and corrections of ill-formed utterances</td>
<td>1. Recasts led to more successful identifications and corrections of their ill-formed utterances. 2. Explicit forms in both types of feedback led to more successful identifications and corrections of their erroneous utterances.</td>
</tr>
<tr>
<td>Yang &amp; Lyster (2010)</td>
<td>72 university students learning English in China</td>
<td>English past tense</td>
<td>Multiple targets</td>
<td>1. Recast group 2. Prompt group 3. Control group 2 dictogloss tasks, 1 question-and-answer activity, 1 picture-cued narrative activity. Approximately 2 hours over a period of 2 weeks</td>
<td>Oral production (story-retelling task), Written production (story composition task)</td>
<td>1. No significant group effect on the oral production of irregular past tense. 2. Significant pre-to-post developments/changes in the accurate use of irregular past tense for all three groups on the oral production (significant pre-to-delayed-post gains for the prompt and control groups, but not the recast group). 3. Only the prompt group made significant pre-to-post and pre-to-delayed-post gains in the accurate use of regular past tense. 4. No significant between-group difference on the written production of irregular past tense. 5. Significant pre-to-post and pre-to-delayed-post improvements for both experimental groups (irregular past tense). 6. No significant group effect on the written production of regular past tense. 7. Significant pre-to-post gains on the written production of regular past tense for all three groups.</td>
</tr>
</tbody>
</table>
8. Significant pre-to-delayed-post gains on the written production of regular past tense only for the prompt group.

Goo (in press) describes 54 Korean EFL learners attending a Korean university in English that-trace filter. Two one-way information gap activities were used: 1. Recast group, 2. Metalinguistic group, and 3. Control group. The experimental groups performed significantly better than the control group on both outcome measures: Grammaticality judgment test and Written production test. No significant differences were found between the recast and metalinguistic groups.

Nassaji's (2009) results were based on a small set of descriptive data (percentage), and no inferential statistically analyses were performed. Goo (in press) also examined the role of WM in the effectiveness of the two corrective feedback, which is later discussed in a relevant section.
Methodological Issues in Empirical Studies on Different Types of Feedback

Their important contributions to the field notwithstanding, most of the experimental and quasi-experimental studies (summarized in Table 1) that investigated the effectiveness of different types of corrective feedback in L2 development are somewhat disputable in terms of modi operandi employed. In what follows, five moot issues, the importance of which has been disregarded in this line of research, are discussed in detail: a) modified output opportunities, b) single-versus-multiple comparisons, c) form-focused instruction, d) prior knowledge, and e) out-of-experiment exposure.

Modified Output Opportunities

First and foremost, modified output opportunities were not controlled for in most of the studies summarized in Table 1; Goo (in press) is the only study that controlled for modified output production. Recasts do not necessarily require learners to produce modified output, but other feedback types (i.e., what has been called prompts) such as clarification requests, metalinguistic feedback, and elicitations, either directly or indirectly, induce modified output. Not controlling for modified output production, which is, in and of itself, not part of corrective feedback, is quite problematic in that modified output production likely leads learners to focus on the linguistic form on which given feedback is intended to provide information. More specifically, as Swain (1995, 2005; see also Swain, 1985, 1993) notes, allowing learners to produce modified/pushed output provides potential opportunities for learners to notice their linguistic problems and test their interlanguage hypotheses in relation to the form in focus, which may contribute to L2 accuracy and fluency. McDonough (2005) showed that modified output opportunities significantly predicted English question development (see Nobuyoshi & Ellis’s (1993) small-scale study for a similar claim). Therefore, it is not too farfetched to postulate that
learners who receive corrective feedback that likely accompanies modified output opportunities (e.g., clarification requests, metalinguistic feedback, elicitations) are in a relatively advantageous position, *ceteris paribus*, compared to those who receive corrective feedback that does not particularly require learners to produce modified output (e.g., recasts) unless modified output production is controlled for.

Accordingly, it is recommended that all learners should either be led to produce output following corrective feedback, regardless of the type of feedback, or be prompted/forced to move on without producing modified output (e.g., Goo, in press; Leeman, 2003; Long et al., 1998). Researchers, who have conducted classroom-based quasi-experimental studies on prompts vs. recasts (e.g. Ammar, 2008; Ammar & Spada, 2006; Lyster, 2004; Yang & Lyster, 2010), tend to argue that controlling for modified output production makes the entire treatment somewhat artificial, and thus ecologically invalid. Pursuing the ecological validity of an empirical study does not necessarily mean the loose control of a potentially confounding variable. In (quasi-) experimental studies, internal validity should precede ecological validity because claims can still be made with respect to independent variables examined in an experimental study even without ecological validity if internal validity is maintained, but no claims can be considered legitimate without internal validity no matter how ecological a given study context may be. Of course, it would be ideal if a study could be designed in a way that satisfies both internal and ecological validity. However, another issue that merits future research is whether modified output can actually change an overall picture of results, that is, how the provision of modified output opportunities can impact the effectiveness of different types of feedback. For instance, prompts with modified output opportunities can be compared to those without production opportunities as well as recasts with and without opportunities, which can provide evidence or counterevidence...
for our criticism vis-à-vis the issue of not controlling for modified output production as well as further evidence for (or some evidence against) the beneficial role of modified/pushed output in L2 learning.

**Comparisons of Single-Versus-Multiple Forms of Feedback**

The relative efficacy of the recast condition on L2 learning has been compared with that of another treatment condition which involves more than one type of feedback (e.g., Ammar, 2008; Ammar & Spada, 2006; Ellis, 2007; Ellis et al., 2006; Lyster, 2004; Lyster & Izquierdo, 2009; Sheen, 2007; Yang & Lyster, 2010). For example, the prompt condition involving multiple types of feedback (e.g., metalinguistic feedback, elicitations, repetitions, etc.) was compared to the recast condition involving only recasts in Ammar (2008), Ammar and Spada (2006), Lyster (2004), and Yang and Lyster (2010); clarification requests were also included in the prompt condition in Lyster’s and Yang and Lyster’s studies. One critical argument against this kind of lop-sided operationalizaion is that because there may be a niche for each type of feedback at which its facilitative role can be maximized, learners receiving multiple types of feedback have better chances of benefitting from contextually appropriate feedback, that is, when its corrective nature is most noticeable, than those exposed to only one type of feedback during the entire task. This is clearly epitomized when Lyster and Izquierdo (2009) utilized repetitions as well as clarification requests. A repetition was used when modified output immediately following a clarification request was still nontargetlike. Therefore, the corrective nature of repetitions used in their study must have become far more salient than when repetitions are used alone as an initial attempt to indicate the ungrammaticality of learner utterances.

Also, focal attention to information delivered through feedback is greatly enhanced by the provision of two different types of feedback one after another in a single turn, which puts
those who receive only one type of feedback on each nontargetlike utterance at a disadvantage (see Ellis, 2007; Ellis et al., 2006; Sheen, 2007). In Ellis et al.’s (2006) study, also reported in Ellis (2007), a combination of a repetition of, and metalinguistic feedback on, a nontargetlike utterance with respect to English past tense, was utilized for their metalinguistic group. Metalinguistic feedback was preceded by a repetition in a single feedback turn by a NS instructor (one of the researchers). Sheen (2007) used both metalinguistic information and recasts for the metalinguistic group (recasts immediately followed by metalinguistic information in a single feedback turn). Ellis (2007), Ellis et al. (2006), and Sheen (2007) all interpreted their results as suggesting metalinguistic feedback is more beneficial to L2 learning than recasts, ignoring the fact that two types of corrective feedback on one erroneous utterance were provided in a single turn according to their operationalization of the metalinguistic treatment, whereas only a single recast was provided for the recast group, not preceded or followed by any other type of feedback. Sheen’s interpretations are even more controversial. Sheen’s metalinguistic group received recasts as well as metalinguistic information in the same turn, that is, the treatment was much more explicit than what metalinguistic feedback/information alone would have brought to learners’ attention. Recasts provided for the metalinguistic group may have played an important role in facilitating the noticing of the corrective nature of recasts as well as targetlike forms contained in recasts due mostly to the metalinguistic information delivered immediately following a recast. That is, the cognitive juxtaposition of a learner’s nontargetlike form and its reformulated version contained in a recast became salient to a great extent by the provision of metalinguistic information. In other words, the performance of the metalinguistic group may have been an outcome of an enhanced combined effect (or a synergy) of recasts and metalinguistic feedback, not metalinguistic feedback alone. There is no evidence that indicates metalinguistic
feedback alone was more effective than recasts in learning English articles in Sheen’s study because there was no metalinguistic feedback only group. To illustrate, Sheen’s research design may be viewed as comparing treatment A as one condition with treatments A + B as another condition, and thus any significant between-group difference indicates the effectiveness of treatment B, but not in comparison with treatment A. It is unreasonable to interpret any between-groups difference as if it were an A-versus-B comparison. Even the impact of treatment B could be overestimated due to an enhanced combined effect caused by the addition of treatment A. Interpreting the between-groups difference as indicating the pure impact of treatment B can be justified only when treatments A and B are not complementary to (or detrimental to) each other, which is not usually the case in research on cognitive behaviors such as language learning.

**Form-Focused Instruction**

An instructional session was also included regardless of the treatment condition as an indispensable part of the experimental treatments in some recast-versus-prompt studies (e.g., Ammar, 2008; Ammar & Spada, 2006; Lyster, 2004; Lyster & Izquierdo, 2009). That is, all learners in the treatment conditions were taught targets at an early stage of the experiment because prompts are assumed to work with linguistic items about which learners have a little knowledge. However, the provision of instructional sessions is a confounding variable/factor that undermines the internal validity of a given study. Lyster (2004) suggested that form-focused instruction may be more effective with prompts than with recasts or no corrective feedback. This indicates that form-focused instruction in fact affected learner performance. In other words, any differential impact of prompts and recasts may not be due entirely to the difference between prompts and recasts in their effectiveness, but due, to some extent, to the moderating role of form-focused instruction. Therefore, Lyster’s overall findings should not be viewed as
suggesting that prompts are more efficacious than recasts; his study provides no (direct) evidence for such a claim. More often than not, an instruction-only group is not even employed (Ammar, 2008; Ammar & Spada, 2006; Lyster & Izquierdo, 2009). Thus, it is by no means clear whether any improved performance witnessed in those studies (without an instruction-only group) is attributable to structured form-focused instruction administered prior to the actual feedback treatment or to corrective feedback provided during the treatment.

As mentioned above, form-focused instruction is provided under the assumption that learners should have basic knowledge of a given target if they are to respond to prompts (Ammar, 2008; Ammar & Spada, 2006; Lyster, 2004; Lyster & Izquierdo, 2009). Nonetheless, given metalinguistic feedback, which is used along with other types of prompts in the prompt condition (e.g., Ammar, 2008; Ammar & Spada, 2006; Lyster, 2004; but see Lyster and Izquierdo (2009) where only clarification requests and repetitions were used for the prompt group), can actually contain/provide relevant grammatical information on any new target(s) that learners have not learned, the rationale for the necessity of administering a form-focused instructional session is not quite convincing. Put differently, it is not unreasonable to compare recasts and prompts in terms of their efficacy on L2 learning without the provision of any formal instruction as long as metalinguistic feedback along with other feedback types constitute the prompt condition. Of course, this kind of comparison without form-focused instruction would be a possible experimental design only if an uneven single-versus-multiple comparison (recasts vs. prompts) discussed earlier could in fact be justified. Strangely, relatively little attention has been paid to the potential of recasts in the acquisition of L2 targets to which learners have never been exposed in this line of research ( prompts vs. recasts).
Prior Knowledge

Prior knowledge was not controlled for in some studies (e.g., Ellis, 2007; Ellis et al., 2006; Loewen & Nabei, 2007). One clear example comes from Ellis (2007). As Ellis mentioned in his discussion on the results, there was room for the metalinguistic feedback group ($M = .689, SD = .265$) to further develop explicit knowledge of the comparative, as shown in the pretest (untimed GJT with ungrammatical sentences), whereas the metalinguistic group already had much explicit knowledge of the past tense morpheme ($M = .844, SD = .108$); that is, any significant pre-to-post development of explicit knowledge of the past tense morpheme was rather unlikely in the first place. However, the recast group already possessed much explicit knowledge of both target morphemes prior to the treatment, as evidenced in their performance on the pretest ($M = .854, SD = .129$ and $M = .855, SD = .159$ for the past tense –ed and the comparative –er, respectively); in other words, it is unsurprising that recasts did not lead to any (differential) effects on the acquisition of the two target forms because there was no room for improvement in the first place, making it highly unlikely to observe any statistically significant treatment effects. Accordingly, the results of Ellis and his colleagues’ studies (Ellis, 2007; Ellis et al., 2006) must be construed in careful consideration of this issue of prior knowledge. Also, in Loewen and Nabei (2007) it was reported that learner accuracy in the untimed grammaticality judgment pretest on English question formation ranged from 70 to 84 percent. Since they conducted an analysis of covariance (ANCOVA) because of the significant between-group difference in learner performance on the pretest, no results were reported on the pre-to-post gains in the untimed grammaticality judgment tests. Obviously, as most learners had already known much about the target before the treatments, it is unsurprising that no significant between-group difference was found in the ANCOVA. But even though the difference had not been statistically
significant in learner performance on the pretest, it is still doubtful that they would have been able to observe any significant pre-to-post gains.

Furthermore, some researchers employed ANCOVAs even there was no covariate that should be taken into account. Ammar and Spada (2006) chose ANCOVAs for all their statistical analyses with learner performance on the pretest as a covariate, none of which seemed to necessitate an ANCOVA, instead of more appropriate mixed ANOVAs (more commonly known as repeated-measures ANOVAs) with Time as a within-subjects variable and Group as a between-subjects variable. In other words, they employed ANCOVAs although they found no significant between-group difference on the pretest (the covariate) and did not even check whether the results of the pretest were correlated with, or predicted, those of the posttests with respect to each relevant dependent variable. Selecting/conducting a separate ANCOVA on learner performance on each separate test administered in each of the two posttest sessions (immediate posttest and delayed posttest) with each relevant pretest as a covariate, in which case the within-subjects variable (i.e., pre-to-post, pre-to-delayed, and post-to-delayed changes) was not analyzed, disburdened their obligation to expound or comment on the improved performance of the control group on the oral tests ($M = 47.5$, $SD = 27.4$ on the pretest, $M = 62.9$, $SD = 19.1$ on the immediate posttest, and $M = 60.9$, $SD = 16.9$ on the delayed posttest), which showed an unexpectedly high increase,\(^2\) and the almost-nonexistent effect of recasts observed in the written tests ($M = 11.1$, $SD = 3.5$ on the pretest, $M = 11.8$, $SD = 3.8$ on the immediate posttest, and $M = 12.3$, $SD = 4.0$ on the delayed posttest), which seems to deserve an explanation given their strong pre-to-post (and pre-to-delayed) improvement in their performance on the oral tests.\(^3\) In addition, Ammar and Spada’s claim that learners with some prior knowledge categorized as high-proficiency learners in the study benefitted more or less equally from recasts and prompts is
rather problematic. They did not find any significant differences among the three groups (in the case of the oral production task) or between the two experimental groups (in the case of the passage correction task) in learner performance on any dependent variable measures. What the results of the oral production measure actually indicate is not that learners with prior knowledge can benefit equally from both types of feedback, but that corrective feedback for those with some prior knowledge in their study may not be as effective as suggested in previous research. Also, the nonsignificant between-group differences on the oral production measure imply that for those with prior knowledge, as long as they participate in interactional activities, learning can take place, for which their study, however, did not provide any evidence due to the absence of a pure control group (a group with no communicative activities). It may be that learners in the control group had enough prior knowledge to take advantage of interactional treatments, even without corrective feedback, as practice opportunities (see their Figure 4 on page 560 in this regard). Thus, prior knowledge functioned as a confound that stranded the chances of distinguishing between the experimental groups and the control group. Also, the nonsignificant difference between the two experimental groups in their performance on the passage correction task in fact resulted from their high pretest scores (see their Figure 3 on page 559). The high-proficiency learners in both treatment groups had already had enough knowledge even before they received experimental treatments. That is, it was unlikely to see any differential effects of recasts and prompts (or any pre-to-post improvement, for that matter) in the first place. Therefore, their claim with respect to high-proficiency learners does not appear to be legitimate.

Another undesirable experimental phenomenon was observed in Yang and Lyster’s (2010) study. They reported somewhat interesting results. Of all three groups (recast, prompt, control), they found the largest pre-to-post effect size for the control group ($d = .94$) on the oral
production of irregular past tense. Also, a large pre-to-post effect size ($d = .82$) was found for the control group on the written production of regular past tense. These excessively high pre-to-post developments by the control group may be attributable, to some extent, to the fact that prior knowledge was not controlled for ($M = 55.56$, $SD = 20.85$ in the oral production pretest on irregular past tense and $M = 67.58$, $SD = 26.36$ in the written production pretest on regular past tense). As mentioned earlier in relation to Ammar and Spada’s study, interactional activities, specifically designed to elicit targets for the experimental purpose, may have evoked learners’ prior knowledge on the targets and provided practice opportunities for learners in the control group even without corrective feedback. Therefore, it may not be too far-fetched a claim, as Long (2007) would also argue, that it is a successful deployment of existing knowledge on the targets, rather than the acquisition of new targets, that actually occurred in their study. It can also be speculated that out-of-experiment exposure to the target forms, to which I turn in what follows, may have contributed to the large pre-to-post developments by the control group.

Regardless of what exactly resulted in the unexpected outcomes, the excessively large pre-to-post developments by the control group appear to jeopardize the validity of their study and, of course, the reliability of their overall findings.

**Out-of-Experiment Exposure**

Lastly, another potential problem ignored in previous research that may limit the validity of previous findings is out-of-experiment exposure to the linguistic form targeted in a given study (e.g., Ellis, 2007; Ellis et al., 2006). For instance, with regard to the metalinguistic group’s improved performance on the oral imitation posttest 2 (with ungrammatical sentences), compared to the oral imitation posttest 1 on the past tense –*ed*, Ellis (2007) suggested that “the effect of the metalinguistic feedback, then, may have been simply that of ‘freshening up’ their
explicit knowledge of this structure, enabling them to attend more closely to the instances of past tense –*ed* in the input they were exposed to between post-test 1 and post-test 2, which were likely to have been plentiful” (p. 359). What this implies is that the participants were exposed to input regarding the past tense –*ed* outside the treatment setting because no treatment was provided between the two posttests, which in turn means the improved performance of the metalinguistic group may not have resulted from metalinguistic feedback that they received during the treatment, calling into question the validity of the overall results. This kind of history effect is a huge threat to the internal validity of empirical studies conducted over a long period of time involving multiple treatment sessions and two or more posttests, especially in a second (as opposed to foreign) language setting. The effect of history is somewhat difficult to control for in toto, nevertheless, should be carefully considered at the initial stage of designing a study so that it can at least be minimized to a great extent.

**Summary of Problems**

Despite their contributions to expanding our understanding of the role of corrective feedback in L2 learning, empirical studies that compared recasts and other types of feedback (e.g., recasts vs. prompts or recasts vs. metalinguistic feedback) fall short of clear and convincing evidence for the claim that metalinguistic feedback or prompts are more effective than recasts due to the lack of methodological soundness. First of all, modified output opportunities were not controlled for in any previous study designed to investigate the relative effect of one type of feedback over another on L2 development. Considering the potential benefits of modified/pushed output (McDonough, 2005; Nobuyoshi & Ellis, 1993; Swain, 1985, 1995, 2005), not controlling for opportunities to produce modified output is quite problematic in an experimental sense because learners receiving feedback that is likely to generate modified output opportunities (e.g.,
prompts) are in a relatively advantageous position, compared to those receiving feedback much less likely to elicit modified output production. Second of all, the beneficial role of recasts has been compared to that of another treatment condition involving two or more feedback types (e.g., Ammar, 2008; Ammar & Spada, 2006; Ellis, 2007; Ellis et al., 2006; Lyster, 2004; Lyster & Izquierdo, 2009; Sheen, 2007). Given each type of feedback may not always function at its maximum in all contextual circumstances, this uneven single-versus-multiple comparison seems unreasonable because the provision of multiple types of feedback is likely to increase the possibility of benefitting L2 development by allowing learners to receive more contextually appropriate feedback in a given feedback episode than when one specific type of feedback is provided. Also, the provision of two feedback types in a single turn (e.g., repetition immediately followed by metalinguistic feedback in Ellis (2007) and Ellis et al. (2006) and recasts immediately followed by metalinguistic feedback in Sheen’s (2007a) study) is much more likely than the provision of one specific type of feedback (e.g., recasts) to facilitate the noticing of corrective feedback, which may contribute to L2 learning. Third of all, form-focused instruction has been included as a part of the experimental treatment in some recast-versus-prompt studies (e.g., Ammar, 2008; Ammar & Spada, 2006; Lyster, 2004; Lyster & Izquierdo, 2009). Form-focused instruction, however, undermines the internal validity of a given study by affecting learner performance, rendering it somewhat unclear whether any improved performance evidenced is attributable to corrective feedback or form-focused instruction. Fourth of all, prior knowledge was not controlled for in some studies so that any significant improvement seemed more or less unlikely in the first place (e.g., Ammar & Spada, 2006; Ellis, 2007; Ellis et al., 2006; Loewen & Nabei, 2007). Also, prior knowledge in combination with interactional activities seems to have contributed to the unexpectedly high pre-post performance of the control group.
Last of all, another potential problem ignored in previous research is out-of-experiment exposure to a linguistic form targeted in a given study (e.g., Ellis, 2007; Ellis et al., 2006), which obviously threatens the internal validity of the study, and thus the reliability of its findings.

However, it should be noted that in addition to these methodological problems, the overall findings of the empirical studies discussed above, which indicate that recasts are not as effective as other forms of feedback (e.g., prompts), may have been attributed to some uncontrolled factors that may have mediated the beneficial role of recasts in L2 learning. I now turn to those potential mediating factors.

**Factors That Mediate the Beneficial Role of Recasts in L2 Learning**

Several factors seem to mediate the degree to which recasts precipitate L2 learning. For example, developmental readiness has been considered as a potential mediating factor. Mackey (1999) and Mackey and Philp (1998) suggested that developmentally more advanced learners could benefit more from recasts than developmentally unready learners. Similarly, researchers have suggested that recasts are more effective for more proficient learners than for less proficient learners (see Ammar, 2008; Ammar & Spada, 2006; Trofimovich et al., 2007). Ammar and Spada (2006) suggested that low-proficiency learners are less likely than high-proficiency learners to benefit from recasts. They further noted that prompts may work better with low-proficiency learners, and that more proficient learners may benefit equally from recasts and prompts. Ammar (2008) also reported that recasts were less effective than prompts with low-proficiency learners based on her data analyzed in terms of developmental stages. Aside from the methodological issues of Ammar’s and Ammar and Spada’s studies, it seems reasonable to assume that proficiency in combination with developmental readiness may affect the efficacy of
recasts on L2 learning. Trofimovich et al. (2007) also found a significant role of proficiency in the effectiveness of recasts in learning lexical and grammatical targets, explaining a statistically significant portion of variance in learner performance. Still, more research on the role of proficiency in the effectiveness of recasts in L2 development needs to be conducted to gather more empirical evidence in this regard. Other potential factors that have been underexplored but may mediate the effects of recasts on L2 learning, which merit much research in the future, include age, gender, interlocutors, settings, and task characteristics (see Mackey, Oliver, & Leeman, 2003 for general discussion on the impacts of age and interlocutors on negative feedback and negotiation patterns; Oliver, 2000 for discussion on the impact of age and settings on the provision and use of negative feedback; Ross-Feldman, 2007 on the relationship between gender and learning opportunities in interaction; Révész, 2009 and Révész & Han, 2006 for discussion on the impact of task characteristics on the effectiveness of recasts).

Also, it has been suggested in previous research that the beneficial effects of recasts depend to a great extent on how they are encoded (e.g., intonation, length, number of changes, etc.), that is, perceptual saliency that affects the noticeability of recasts (e.g., Egi, 2007b; Loewen & Philp, 2006; Philp, 2003; Sheen, 2006). In general, short/partial recasts with one or two corrective changes are more effective at facilitating noticing than long/full recasts with more changes. It should be noted, however, that this saliency/noticeability of recasts, as discussed in Nicholas et al. (2001), may also be affected by research contexts. Nicholas et al. (2001) suggested that mixed results of L2 studies on recasts may be attributable to contextual differences with laboratory studies more likely to produce positive effects of recasts on L2 learning than classroom-based research because learners may become more attentive to recasts in tightly-controlled lab-based studies, compared to those participating in classroom-based ones.
(see Mackey & Goo, 2007 for a meta-analytic review of interaction research in terms of classroom vs. laboratory research settings).

The type of target structure may also mediate the role that recasts play in L2 learning. That is, recasts may work with some linguistic areas or some grammatical structures, but not with others (e.g., Iwashita, 2003; Leeman, 2003; Mackey et al., 2000; Ortega & Long, 1997). For instance, in relation to learners’ perceptions about recasts, Mackey et al. (2000) suggested that recasts may not be an optimal choice for the provision of morphosyntactic feedback, and that the learners’ less accurate perception of morphosyntactic feedback evidenced in their study may have been caused by a communicative need that led the learners to prioritize understanding rather than the accuracy of such morphosyntactic features as agreement and plural formation unlikely to result in a communication breakdown, unlike phonology and lexis likely to undermine their understanding and interaction as a whole. Consequently, much of the morphosyntactic feedback provided may have gone unnoticed due to the paucity of an appropriate level of attention. Of course, there is a variation even within a linguistic category. Ortega and Long (1997, also in Long et al., 1998) found that the recast group outperformed the model group in the acquisition of Spanish adverb placement, but not direct object topicalization (no learning took place under either treatment condition.) although both structures are considered at the same developmental stage. Iwashita (2003) also showed that recasts had a significantly positive impact on the acquisition of Japanese te-fom verb, but were not effective in learning locative word order or locative particle use. Also, in Leeman’s (2003) study, the beneficial effects of recasts on the development of Spanish gender agreement evidenced on the immediate posttest were not found on the second posttest, resulting in a nonsignificant difference between
the recast group and the control group. But for Spanish number agreement, the recast group performed significantly better than the control group on both test sessions.

Research findings are still inconclusive and suggest that perceptual salience of different target forms/structures may differentially affect the extent to which learners can benefit from recasts. Long (2007) suggested that unobtrusive implicit negative feedback (e.g., recasts) works for salient linguistic features and more explicit types of feedback (e.g., metalinguistic feedback) works for communicatively redundant and phonologically nonsalient linguistic items (e.g., past tense -ed). He also noted that recasts may be more effective for linguistic structures/forms that are hard to learn and thus require long-term treatments, compared to easy structures/forms that short-term treatments of explicit feedback may be sufficient to learn.

Also, individual differences in such areas as language aptitude (see Sheen, 2007a in this regard), cognitive abilities, intelligence, personality, motivation, learning styles, and learning strategies (see Dörnyei, 2005 for detailed discussion on individual differences in SLA or Dörnyei, 2006 for a summary of his discussion) may delimit the effectiveness of recasts that occur during interaction. Especially, considering previous research that has indicated a potential role of WMC in L2 learning (e.g., Geva & Ryan, 1993; Goo, 2010, in press; Harrington & Sawyer, 1992; Havik et al, 2009; Kormos & Sáfár, 2008; Mackey et al., 2002, 2010; Mackey & Sachs, in press; Miyake & Friedman, 1998; Rai et al., 2011; Robinson, 2005b; Sagarra, 2007, 2008; Sagarra & Herschensohn, 2010; Trude & Tokowicz, 2011; Tokowicz et al., 2004; Walter, 2004; see also Juffs & Harrington, 2011; Williams, in press for reviews of research on WM and L2 learning), individual differences in WMC are likely to mediate beneficial effects of recasts. That is, learners with high WMC may notice more recasts (Makey et al., 2002) and benefit more from recasts than those with low WMC (Goo, in press; Sagarra, 2007). Interestingly,
Trofimovich et al. (2007) did not find any significant role of WMC in the efficacy of recasts on the acquisition of English possessive determiners, but observed that attention control, phonological memory, and language aptitude (analytical ability) were related to the effectiveness of recasts in learning the grammatical target. Especially, attention control, *inter alia*, was found to explain a significant amount of variance in learner performance in all three areas: lexical, grammatical, and mixed (lexical + grammatical) areas. However, only a few studies have attempted to investigate whether and how WMC affects the efficacy of corrective feedback on the actual development of L2 linguistic forms (Goo, in press; Mackey et al., 2002; Mackey & Sachs, in press; Sagarra, 2007; Trofimovich et al., 2007). Among them, only one empirical study has ever explored the relationship between WMC and the effects of different types of corrective feedback on L2 learning (Goo, in press). Given this paucity of relevant interaction research (and inconsistent findings, for instance, Trofimovich et al.’s finding of no significant relationship between WMC and recasts, as opposed to those that indicated a nontrivial association between them), much more research on the impacts of WMC on the role of recasts in L2 learning needs to be conducted to get more clear-cut evidence of the relationship between WMC and recasts. Of course, it is also worth exploring how WMC affects positive effects of corrective feedback on the acquisition of different target structures. Discussed in what follows is relevant research on the nature of WM and individual variation in WMC and its relation to performance on higher-order cognitive tasks including language learning.
Chapter II: Working Memory Capacity (WMC)

Overview

Since the inception of the concept of WM (e.g., Baddeley & Hitch, 1974) that began to replace that of short-term memory (STM generally defined as a system for the temporary storage of information), WM has been given a position of pivotal importance in the field of cognitive psychology (and other research areas that deal with human cognition), generating a barrage of theoretical explanations and numerous empirical studies with respect to the nature of WM and its relation to higher-order cognitive behaviors such as reading, reasoning, general fluid intelligence, language comprehension, and so forth (see Baddeley, 2007; Conway et al., 2007a; Jarrold & Towse, 2006; Miyake, 2001; Miyake & Shah, 1999 for varied theories and models about WM and reviews of relevant theoretical developments). In general, WM is viewed as “the ability to maintain information in an active and readily accessible state, while concurrently and selectively processing new information” (Conway, Jarrold, Kane, Miyake, & Towse, 2007b, p. 3). Similarly, Jarrold and Towse (2006) defined it as “the ability to hold in mind information in the face of potentially interfering distraction in order to guide behavior” (p. 39). That is, WM comprises mechanisms dedicated to active maintenance of information and mechanisms for cognitive control that coordinate and integrate its storage and processing operations to guide task-relevant behaviors. Due to these mechanisms, WM is considered as a broader concept than STM and thought to be involved in a wide range of higher-order cognitive performance, serving a critical function in human cognition. Hence, individual differences in WM capacity may well be reflected in differences in performance on varied complex cognitive tasks such as language comprehension including L1 reading and sentence processing (e.g., Daneman & Carpenter, 1980, 1983; Just & Carpenter, 1992; King & Just, 1991; MacDonald, Just, & Carpenter, 1992; see
Daneman & Merikle, 1996 for a meta-analysis of the research findings on the relationship between language comprehension and WM capacity, $rs = .41$ and .52 for global comprehension and specific comprehension, respectively) as well as general intellectual abilities including reasoning and general fluid intelligence (e.g., Conway, Cowan, Bunting, Therriault, & Minkoff, 2002; Conway, Kane, & Engle, 2003; Engle, 2002; Engle, Tuholski, Laughlin, & Conway, 1999; Kane et al., 2004; Süß, Oberauer, Wittmann, Wilhelm, & Schulze, 2002; see also Ackerman, Beier, & Boyle, 2005 for a meta-analysis of the findings from studies on the association between WM capacity and general fluid intelligence, in which the estimated population correlation coefficient = .479 and Kane, Hambrick, & Conway’s (2005) response to Ackerman et al.’s meta-analysis based on their reanalysis that found $r = .72$). However, one interesting claim, which is somewhat counterintuitive, emerged from a category learning experiment; sometimes less may be more (DeCaro, Thomas, & Beilock, 2008). DeCaro et al. (2008) found that high WM participants were faster in learning rule-based categories, but slower in learning information-integration categories, than their low-span counterparts, suggesting sometimes less WM may be better than more WM. These findings, they argued, may have been attributed to low-span individuals’ use of procedural-based strategies due to their inability to employ explicit learning strategies reliant on executive control. However, DeCaro, Carlson, Thomas, and Beilock (2009), in a replication of DeCaro et al.’s (2008) study, showed that low-WM individuals in fact employed simple rule-based strategies to perform information-integration tasks, but not procedural-based learning strategies. They suggested that low WM individuals’ reliance on simple rule-based strategies may have contributed to their better performance, compared to the performance of their high WM counterparts, on the information-integration task observed in DeCaro et al. (2008).
Research has shown that WM and STM are separable constructs, but strongly correlated with each other, and that the most vital aspect of WM lies in its functional superiority to STM in predicting performance on complex cognitive tasks (Conway et al., 2002; Daneman & Merikle, 1996; Engle, Tuholski et al., 1999). For example, Daneman and Merikle (1996) in their meta-analysis showed that comprehension was correlated more strongly with verbal WM measures ($r_s = .41$ and $.52$ with the global and specific measures of comprehension, respectively) than with STM measures ($r_s = .28$ and $.40$ with the global and specific comprehension), confirming Daneman and Carpenter’s (1980) findings ($r_s = .72$, $.90$, and $.59$ between the reading span and comprehension on facts and pronoun references, as well as between the reading span and verbal SAT scores, respectively, as compared to $r_s = .37$, $.33$, and $.35$ between the word span test and comprehension). Moreover, in their latent variable analysis Engle, Tuholski et al. (1999) found that after the common component to WM and STM was factored out, the residual WM variance was still significantly correlated with general fluid intelligence ($r = .49$), but that the correlation between the residual STM variance and general fluid intelligence was not significant ($r = .12$) after that common component was removed. Similarly, Conway et al.’s (2002) latent variable analysis yielded evidence that WM capacity is a better predictor of performance on general fluid intelligence ($r = .60$) than is STM capacity ($r = .18$).

**Nature of WM and WMC variation**

Although there appears to be a consensus regarding the predictive power of WMC for higher-order cognitive performance, the issues of the exact nature of WM and of variation in WMC (e.g., how does WM work? and what are the critical sources of WM variation?) are controversial and inconclusive among cognitive psychologists. According to the resource-sharing account (Daneman & Carpenter, 1980, 1983), WM capacity is a limited pool of cognitive
resources and the amount of information that can be stored during processing depends on how efficiently such processing can take place, a trade-off between the processing and storage demands. That is, individuals with high WMC are able to process a given concurrent task more efficiently, making available more cognitive resources for storage than those with low WMC. In a variant of the resource-sharing account (the capacity theory), Just and Carpenter (1992) argued that “the trading relation between storage and processing occurs under an allocation scheme that takes effect when the activation maximum is about to be exceeded” (p. 123), noting that “performance differences among individuals will emerge, primarily when the task demands consume sufficient capacity to exhaust some subjects’ resources” (p. 143).

The task-switching account is an alternative proposal to the resource-sharing account about the nature of WM and variation in WMC (Towse & Hitch, 1995, 2007; Towse, Hitch, & Hutton, 1998). According to this account, WMC is limited because individuals undergo the rapid forgetting of to-be-remembered items during the time spent processing. Thus, performance on WM span tasks is to a large extent determined by the temporal dynamics of WM span, which points to the intrinsic involvement of processing efficiency and time in the maintenance or loss of temporary information. In WM span tasks, individuals are required to switch back and forth between the processing and storage components of the task. The longer they are engaged in the processing component of the task, the more susceptible to-be-remembered information becomes to forgetting, which implies that processing efficiency does have an impact on WM task performance. The task-switching account differs from the traditional resource-sharing account in that processing and storage function independently of each other.

Barrouillet, Bernardin, and Camos (2004; see also Liefooghe, Barrouillet, Vandierendonck, & Camos, 2008) have proposed a somewhat variant WM model, a combination
of limited resource-sharing and task-switching accounts. Their so-called time-based resource-sharing model suggests task concurrent maintenance of information in WM is impaired by the amount of time for attentional switching. They noted “the crucial element is the amount of time attention is occupied by task switching and not the amount of information to be maintained actively during task switching” (Liefooghe et al., 2008, p. 491). Their time-based resource-sharing model differs from the traditional resource sharing account in that it assumes no impact of memory load on task switching (processing), and at the same time is also distinguished from the task switching account as it assumes that the attention-switching process is resource-limited, and thus considers the difficulty of processing operations; the task-switching account is concerned only with the duration of processing regardless of the difficulty of processing.

However, some studies on processing efficiency in terms of how processing time (Engle, Cantor, & Carullo, 1992), difficulty (Conway & Engle, 1996), or processing speed (Conway et al., 2002) affects the association between WM capacity and cognitive behaviors (e.g., comprehension or general fluid intelligence) have failed to yield any significant evidence that processing efficiency actually has an impact on the correlation between WM capacity and cognitive performance, at least within an age group (young adults). Friedman and Miyake (2004) also found, in their analyses on the data from the experimenter-administered condition, no significant change in the span-comprehension correlations after both processing times and word span scores were partialled out, compared to the correlations when only word span scores were partialled out, providing further evidence against predominantly processing-based accounts of span differences (e.g., task-switching). In addition, the results showed that sentence position slopes assumed to reflect the effect of memory loads on the reading span performance did not
mediate the span-comprehension correlations in any significant way, as evidenced in their partial correlation analyses, rejecting the resource-sharing account as well.

Another line of theoretical explanation regarding the nature of WM and individual differences in WMC is the executive attention or controlled-attention view which has been developed by Engle and his colleagues (Engle, 2002; Engle, Kane, & Tuholski, 1999; Engle, Tuholski et al, 1999; Kane, Bleckley, Conway, & Engle, 2001; Kane, Conway, Hambrick, & Engle, 2007). According to the executive attention view, individual differences in WMC that lead to performance differences in complex cognitive tasks derive mainly from variation in domain-general executive attention processes and, to some extent, from variation in domain-specific storage and rehearsal processes. Kane et al. (2004) found strong correlations between verbal WMC and spatial WMC latent variables, which epitomizes this domain generality of WM construct and the executive attention ($r_s = .93$, .84, and .83 in three different path models). In other words, WMC span tasks capture primarily the executive attention, and secondarily the domain-specific storage and rehearsal capacity, whereas STM tasks reflect primarily the domain-specific storage and rehearsal capacity, and secondarily the domain-general executive attention. That is, according to them, no span task is a completely pure measure of either WMC or STM capacity (Conway et al., 2002; Engle, Kane et al., 1999; Engle, Tuholski et al., 1999; Kane et al., 2007). This view implies that WM and STM are not mutually exclusive, but strongly correlated with each other, as mirrored in nontrivial high correlations between STM tasks and WMC tasks observed in previous latent variable studies (e.g., $r_s = .68$ and .82 in Engle, Tuholski et al., 1999 and Conway et al., 2002, respectively; $r_s = .79$ and .89 between the verbal WMC and verbal STM latent variables and between the spatial WMC and spatial STM, respectively, in Kane et al., 2004). Engle and his colleagues in their recent proposal (e.g., Kane et al., 2007) suggested
that one fundamental strength of high-WMC individuals is their superior use of executive control, vis-à-vis low-WMC individuals, in simple attention-control tasks involving no memory components such as dichotic listening\(^4\) (Conway, Cowan, & Bunting, 2001), Stroop tasks\(^5\) (Kane & Engle, 2003), and antisaccade tasks\(^6\) (Kane et al., 2001), all of which showed high-span participants’ advantages over low-span participants in maintaining task goals and controlling attention accordingly. They went so far as to claim that “a third variable, representing a low-level executive attention capability, influences functioning on all of these selective-attention, WM-span, and memory-retrieval tasks (and, presumably, on indices of Gf as well)” (Kane et al., 2007, p. 34). Their executive attention theory also points to high-span individuals’ superior ability to suppress attention to irrelevant tasks and the activation of task-irrelevant information (Kane & Engle, 2000; Rosen & Engle, 1997, 1998).

Similarly to the executive attention view, Hasher and her colleagues (Hasher, Lustig, & Zacks, 2007; Hasher & Zacks, 1988; Hasher, Zacks, & May, 1999) contend that inhibitory-based executive control is an essential factor that determines the scale and scope of the predictive power of WM tasks, noting that “inhibitory efficiency accounts for much of the variation in cognitive performance...inhibition is a fundamental determinant of the apparent differences in what many investigators term ‘working memory capacity’” (Hasher et al., 2007, p. 229).

Somewhat differently from the executive attention view, however, the inhibitory control account is concerned predominantly with the suppression of goal-irrelevant information rather than with active maintain of information in the face of interference. According to this view, inhibitory control functions “in the service of goals to (1) prevent irrelevant information from gaining access to the focus of attention, (2) delete [italics in original] no-longer relevant items from consideration, and (3) restrain [italics in original] prepotent responses so that other, initially
weaker response candidates can be evaluated and influence behavior as appropriate for current goals” (Hasher et al., 2007, pp. 230-231). Hasher and her colleagues further argue that inhibitory control mechanisms are key sources of much of age-related and intra-individual variation in WMC and in complex cognitive tasks (Hasher et al., 2007; see also Lustig, May, & Hasher, 2001 for evidence of the effects of interference on age-related variation in WMC and on the correlation between WM span and cognitive performance).

From a similar perspective of cognitive control, Braver, Gray, and Burgess (2007) suggest that proactive control mechanisms serve to prevent interference, while reactive control processes can detect and suppress interference when it occurs. Braver et al. view controlled processing, which they consider as the crux of working memory, as an emergent phenomenon that arises from dynamic interactions among relevant processing subsystems in the brain. They explain variation in working memory in terms of the critical role of the prefrontal cortex (PFC) in cognitive control, and two distinct modes of cognitive control, proactive and reactive control. Implying that cognitive mechanisms such as executive attention and inhibitory control are highly involved in individual variation in WMC, Unsworth (2007) argued that one pivotal difference between high- and low-WMC individuals lies in the size of search set of to-be-recalled items with low-WMC individuals utilizing a larger search set (a combination of present and previous target items) than do high-WMC individuals, based on his finding that low-WMC individuals made significantly more errors of previous list intrusions (items from previous lists) than did high-WMC individuals. Similarly, Unsworth and Engle (2007) claimed that low-WMC individuals are more vulnerable to proactive interference (PI) than high-WMC individuals because low-WMC individuals are poor at delimiting their search sets to task-relevant representations, rendering a retrieval process quite challenging. They further note, using the
concepts of primary memory and secondary memory, that variation in WMC may be driven by individual differences in the ability to maintain information and task goals accessible in primary memory (when active maintenance of information is required) and the ability to retrieve task-relevant information from secondary memory in the presence of competition.

However, Jarrold and Bayliss (2007) consider storage constraints as another important culprit of individual and developmental variation in WM performance, arguing variation in WMC stems from three independent sources: storage capacity, processing efficiency, and executive ability to coordinate or combine the two demands. Various other theoretical explanations and models have been proposed in strenuous efforts to understand the nature of WM and sources of individual variation in WMC (see Conway et al., 2007a; Miyake & Shah, 1999 for diverse theoretical models). Based on their research on expert performance, Ericsson and his colleagues (Ericsson & Delaney, 1999; Ericsson & Kintsch, 1995) proposed the existence of long-term WM. Under this view, through the acquisition of domain-specific skills, individuals are able to encode relevant information into long-term memory in a readily accessible form so that the encoded information can be rapidly retrieved from long-term memory whenever needed during domain-specific activities. Another interesting proposal is that there may be a specialized system in the central executive dedicated predominantly to on-line syntactic processing and other related aspects of language processing, a separate WM system that is not represented by general verbal WM as measured by standard WM tests (Caplan, Waters, & DeDe, 2007). In their connectionist-based perspective on language comprehension, MacDonald and Christiansen (2002) contended that the reading span task is no more than a measure of language processing skills, for that matter, performance on the reading span task does not reflect an independent WM function. They further noted that individual differences in language processing
skills including reading span tasks (according to their argument) stem from two factors and interaction between them, variations in exposure to language and biological differences (e.g., precision of phonological representations). Nevertheless, it would be hard to imagine the predictability of other nonlinguistic verbal WM span measures such as operation span tasks for reading comprehension (e.g., Engle, Nations, & Cantor, 1990; Turner & Engle, 1989; see also Daneman & Merikle, 1996 for more information in this regard) and significant correlations, as evidenced in previous studies, between WMC with processing factors partialled out and cognitive performance (e.g., Conway & Engle, 1996; Conway et al., 2002; Engle et al., 1992), without assuming the existence of a domain-general WMC that underlies varied complex cognitive behaviors, one of which is language comprehension.

**WMC and Language Comprehension**

As mentioned above, it has been widely accepted that WMC relates to language comprehension (e.g., Daneman & Carpenter, 1980, 1983; Just & Carpenter, 1992; King & Just, 1991; MacDonald, Just, & Carpenter, 1992; Swanson & Berninger, 1995; Turner & Engle, 1989; see Daneman & Merikle, 1996 for a meta-analysis). Daneman and Carpenter (1980) found that WMC (reading span) was significantly correlated with three reading comprehension measures ($r = .59$, $.72$, and $.90$ for the correlations with performance on verbal SAT, fact questions, and pronouns reference questions, respectively); the correlations were also significant between their listening version of reading span task and reading comprehension ($r = .53$, $.67$, and $.72$ for the correlations with verbal SAT, fact questions, and pronoun reference questions, respectively). Daneman and Merikle’s (1996) meta-analysis, based on the data from 6,179 participants in 77 studies, yielded further evidence that WMC is a good predictor of language comprehension ($r = .41$ and $.52$ for the correlations, in the case of WM span measures involving a verbal
processing component such as reading and listening span tasks, with global comprehension and specific comprehension, respectively). Researchers have also made continuous efforts to investigate the role that WM plays in other areas of language acquisition. For example, Daneman and Green (1986) showed the importance of WMC in vocabulary learning, especially in drawing inferences about the meanings of novel words from contextual cues. Evidence for the predictive power of WMC for language acquisition also derives from numerous studies, accounting for individual differences in verbal fluency (Daneman, 1991) and the ability to deal with syntactic ambiguity and to process syntactically complex sentences (Just & Carpenter, 1992; King & Just, 1991). King and Just (1991), for instance, showed that although lower-WMC individuals spent more time processing syntactically complex sentences (object relative clauses), their performance on comprehension was poorer than that of higher-WMC individuals, but that no span-differences were evidenced in reading time for subject-relative sentences, undemanding sentences that consume much less WM resources, exemplifying the selective effects of WM constraints on syntactic processing. Also, it has been suggested that phonological short-term memory (PSTM) or phonological loop is closely associated with vocabulary development (e.g., Baddeley, Gathercole, & Papagno, 1998; Gathercole & Baddeley, 1989). Baddeley et al. (1998) noted that the critical function of the phonological loop is not simply to store phonological representations of familiar words, but to facilitate the learning of new vocabulary. Similarly in SLA research, the phonological loop in Baddeley’s multi-component model of WM has been deemed to function as an important determinant of L2 learning (e.g., Ellis, 1996; Williams & Lovatt, 2003), apart from rather unitary perspectives of WM emphasizing mostly on a domain-general executive control of WM. Reviewed in the following section is previous SLA research on whether and how individual differences in WMC affect L2 learning.
WMC and SLA

Reflecting a burgeoning interest in research on individual differences in terms of how they affect the scale and scope of L2 development, WMC has been discussed as a potentially critical component of language aptitude in the field of SLA (e.g., Miyake & Friedman, 1998; Robinson, 2005a; Skehan, 2002; see also Juffs & Harrington, 2011; Williams, in press for reviews on WM and L2 learning) and investigated with respect to whether and how individual differences in WMC affect L2 learning in various research areas such as reading comprehension (e.g., Geva & Ryan, 1993; Harrington & Sawyer, 1992; Rai et al., 2011; Walter, 2004), incidental L2 rule learning (Robinson, 2002, 2005b), processing of redundant grammatical forms (Sagarra, 2008), syntactic processing (e.g., Havik et al., 2009; Juffs, 2004, 2005; Miyake & Friedman, 1998), L2 pronunciation (Trude & Tokowic, 2011), overall L2 proficiency as well as four major skills (e.g., Kormos & Sáfár, 2008), potential impacts of think-alouds on L2 performance (Goo, 2010), L2 processing in relation to proficiency (Sagarra & Herschensohn, 2010), L1-L2 translation strategies of learners with study-abroad experiences (Tokowicz et al., 2004), L2 grammar learning through corrective feedback (e.g., Goo, in press; Mackey & Sachs, in press; Mackey et al., 2002; Trofimovich et al., 2007; Sagarra, 2007), and modified output production (e.g., Mackey et al., 2010; Sagarra, 2007). Described below is a summary of L2 research that has investigated the role of WM in L2 learning.

Harrington and Sawyer (1992), one of the earliest studies on WM and L2 learning, examined how sensitive L2 WMC was to differences in L2 reading performance among advanced Japanese ESL learners. Digit span, word span, and reading span measures in both L1 and L2 versions were developed and utilized. The results showed that L2 reading span scores were significantly correlated with two TOEFL measures (Grammar and Reading comprehension,
Neither L2 word span nor L2 digit span was significantly correlated with TOEFL reading, resonating with Daneman and Merikle’s (1996) finding, but in L2 contexts, that complex span measures such as reading span and operation span tasks designed to tap both storage and processing abilities are better predictors of reading comprehension than simple span measures such as word span and digit span tasks designed to tap storage only. The correlation between the two versions of reading span tasks (L1 and L2 versions) was moderate, but statistically significant ($r = .39, p < .05$); interestingly, no correlation coefficients were reported between L1 reading span and TOEFL measures.

Geva and Ryan’s (1993) study, where 73 Grade 5-7 students attending an English-Hebrew bilingual school participated in a series of cognitive and linguistic tasks, initially found significant correlations between L1 WM measures and L1 linguistic skills. After effects of intelligence and grade level were partialled out, however, the initially observed significant correlations dropped. In some cases, they lost significance. Nevertheless, L2 working memory measures were significantly correlated with L2 linguistic tasks, and their statistical significance was maintained even after intelligence and grade level were factored out. Also, L1 WM was correlated with performance on L2 linguistic tasks, and L2 WM was correlated with L1 linguistic skills. However, when intelligence and grade level were partialed out, the observed crosslinguistic correlations significantly dropped and lost statistical significance. In addition, they showed L2 WM made a significantly unique contribution to L2 reading comprehension.

Miyake and Friedman (1998) observed that L2 WMC measured via a listening span task was significantly correlated with L2 syntactic comprehension. In their path analysis, Miyake and Friedman showed the significant contributions of L2 WMC (listening span) to L2 syntactic
comprehension with a path correlation coefficient of .38 ($p < .01$). The correlation between L1 and L2 listening span tasks was statistically significant ($r = .58, p < .01$), which parallels previous findings that showed significant L1-L2 WM correlations (e.g., Mackey et al., 2002; Osaka & Osaka, 1992; Osaka, Osaka, & Groner, 1993; see also van den Noort, Bosch, & Hugdahl, 2006 for correlations involving L3 as well as L2).

Robinson (2002, 2005b) reported statistically significant correlations between WMC (measured via a L1 reading span) and learner performance on immediate and delayed listening grammaticality judgment posttests ($rs = .42$ and $.48$, respectively, $p < .01$ for both) as well as on two delayed production tests ($rs = .33$ and $.44$, respectively, $p < .05$ for both), providing some evidence that WMC may affect incidental learning of L2 rules (Samoan in his study). Interestingly enough, neither of the other two experimental conditions (explicit series-solution learning condition and implicit artificial grammar learning condition) was significantly correlated with WMC ($rs = .04$ and $.09$, correlations between WMC and the explicit condition and between WMC and the implicit condition, respectively).

Regarding the effects of WMC and study-abroad experience on error types during single-word translation, Tokowicz et al. (2004) found that higher-WMC learners with more study-abroad experience tended to make efforts to communicate utilizing approximate translations even when they were unable to provide exact translations whereas lower-WMC learners tended to make non-response errors (no response when they did not know) regardless of their study-abroad experience. They suggested that higher-WMC learners were able to use this strategy due to their ability to maintain multiple items simultaneously in the absence of correct translations, indicating the possible role of WMC in developing communicative strategies.
Walter (2004), using two groups of French learners of English (lower- and upper-intermediate learners of English), investigated whether/how verbal WM was linked to reading comprehension (summary completion scores) and structure-building skills (pro-form resolution scores) in both L1 and L2. WM was measured via the learners’ performance on L1 and L2 reading span tasks. Walter found a significant overall correlation between WM and reading comprehension in both L1 and L2. However, separate correlation analyses showed no significant correlation between L1 WM and L1 reading comprehension for either group. As regards L2 WM and L2 reading comprehension, a significant correlation was observed in the performance of the lower-intermediate group, but not that of the upper-intermediate group. With regard to learner performance on the structure-building task (remote pro-form resolution), although combined WM scores (in L1 and L2 WM) were significantly correlated with the performance of the low-intermediate group, when only L2 WM was considered, no significant correlation was observed. Likewise, L2 WM was not significantly correlated with the performance of the upper-intermediate group on the L2 structure-building task.

In a Hungarian EFL context, Kormos and Sáfár (2008) showed that WMC measured via a backward digit span task was significantly correlated with the improved performance of the beginner group on three of four major skills (reading, listening, and speaking, but not writing) and the use of English (knowledge of vocabulary and grammatical constructions) as well as with the overall proficiency scores. WM accounted for 30.25% of the total variance in the performance of the beginner group on the complete language test. In addition, their results showed that phonological short-term memory and working memory may be different entities and contributed to L2 learning in different ways. No correlations were found between language skills
and PSTM for beginners. However, for pre-intermediate learners, PSTM was found to be correlated with L2 learning.

Mackey et al. (2002) investigated how individual differences in WM are related to the noticing of interactional feedback (i.e., recasts) and L2 English question development. WM was measured via three memory tasks (nonword recall, L1 and L2 listening span tasks); they used composite scores. Thirty adult Japanese ESL learners carried out three interactional treatment tasks (picture drawing, picture difference, and story completion) and two similar assessment tasks (picture drawing and story completion). The study found that more noticing occurred among high-span learners, whereas less noticing was associated with low-span learners. Interestingly, the beneficial effects of recasts on English question development by high-span learners were evidenced in the delayed posttest, but not in the immediate posttest in which more development seemed to occur among low-span learners, which appears somewhat counterintuitive. Mackey and Sachs (in press) recently conducted a similar study, but this time, with 9 aged Spanish-speaking learners of English (mean age: 72). WM measured via an L1 listening span task was found to be significantly associated with the development of English question forms evidenced in learner performance on the immediate posttest. The findings of Mackey and Sachs (in press) and Mackey et al. (2002), albeit interesting, were based on quite limited sets of data, and thus, as they cautioned, must be interpreted with care. Nevertheless, both studies imply a potential role of WM in interaction-driven L2 learning.

Potential effects of WMC on noticing was confirmed in Mackey et al.’s (2010) study in which forty-two English-speaking learners of Spanish as a foreign language participated in four interactional tasks (a map task, a picture drawing task, a spot-the-difference task, and a story completion task) with a Spanish NS in dyads. A thirty-six-item L1 listening span was
administered to their participants. They found that WMC significantly positively predicted the production of modified output ($\beta = .41$ between the composite WM scores (processing + recall) and modified output production, and $\beta = .31$ between their recall scores and modified output production). Another interesting finding is that the processing part of the WM test was related to the production of modified output by making changes to original nontargetlike utterances, but not by repeating recasts. It should be noted, however, that because they focused on modified output production, most feedback was provided via clarification requests and repetitions, not in the form of recasts, that is, this issue of the relationship between the processing component of WM and leaners’ tendency to produce modified output by changing their utterances definitely merits future research. If modified output production reflects noticing, albeit not entirely nor always, and if in fact noticing is a *sine qua non* for L2 development to occur, as Schmidt argues (1990, 1995, 2001), Mackey et al.’s findings suggest that high-WMC learners may benefit more from interactional feedback and modified output than low-WMC learners.

With regard to recasts, Sagarra (2007) examined the effectiveness of computer-delivered oral recasts and the role of WM in learning Spanish gender and number agreement. She found that WM measured via a reading span task was statistically significantly related to the performance of the recast group on both written and oral face-to-face interaction tasks, but not to that of the control group in either measure. WM was also found to be associated with the amount of targetlike modified output for both gender and number agreement. However, given each recast was followed by a pause designed to provide an opportunity for the learner to produce modified output in the face-to-face interaction test, the finding on the relationship between WM and targetlike modified output should not be overgeneralized as evidence of the direct link between WM and modified output production.
Sagarra (2008) also investigated how WM is related to L2 processing of redundant grammatical features in relation to adverb-verb tense agreement (e.g., a preterit verb with a past temporal adverb). One hundred and fifty-six adult English-speaking learners of Spanish participated in a non-cumulative moving window task and a L1 reading span task (in addition to three screening tasks). Sagarra found a significant effect of WM and of the WM x Agreement (agreement/disagreement) interaction, that is, WM affected low-proficient L2 learners’ ability to process these redundant grammatical features when they needed to attend to meaning simultaneously.

More recently, Sagarra and Herschensohn (2010) compared two Spanish FL proficiency groups (intermediate and beginners) and a Spanish monolingual group in terms of reaction time at adjectives in a moving window task, accuracy in a grammaticality judgment task, and the relationship between WM and their performance. WM was measured via a reading span task. Results showed that high-WMC intermediate learners were more sensitive to gender agreement violations than their low-WMC counterparts, which they claimed suggests it is more cognitively demanding to process/notice gender agreement violations than number agreement violations. However, WM was not correlated with the performance of the beginner group due to its overall low WMC. As for the Spanish monolingual group, ceiling effects precluded any significant correlations.

Havik et al. (2009) examined the potential association between WM and L2 learners’ online sentence processing of Dutch relative clause constructions. In a moving window task, German advanced learners of Dutch read sentences containing ambiguous Dutch subject and object relative clauses at their own pace. WM was measured in two versions of Daneman and Carpenter’s (1980) reading span task; their L1 (German) and L2 (Dutch) versions. Results from
Experiment 1 showed that high-WM L2 learners displayed a processing advantage for subject over object relative clauses in the items with a short distance between the NP modified and the point of disambiguation, which is a similar pattern observed in the low-WM native speaker group. However, WM did not affect the L2 learners’ processing in Experiment 2 where they were not required to perform a semantic verification immediately after each sentence, unlike Experiment 1.

Rai et al. (2011) conducted an interesting study in which WM was examined in relation to stress and L2 learners’ ability to draw inferences during reading comprehension. Fifty-five English-speaking learners of Spanish as a foreign language carried out a series of questionnaire measures, a reading comprehension task, and a WM task (operations span). The researchers found that learners with higher WM obtained higher accuracy across all three levels of inferences than those with lower WM: non-inference (facts), bridging inference (pronominal connections), and pragmatic inference (inference using world knowledge). Also observed was, albeit nonsignificant, that learners with higher WM tended to be somewhat slower in processing pragmatic inference questions (also non-inference questions) than those with lower WM. In addition, learners with lower WM were found to be more susceptible to stress, compared to those with higher WM, in terms of processing efficiency (a much larger increase in reaction time). They interpreted their findings as indicating higher-WM learners’ general tendency to trade processing efficiency (slower processing time) for comprehension accuracy, which also seems to be the case with lower-WM learners only under high stress although less successful.

Trude and Tokowic (2011) explored how WMC was associated with L2 pronunciation learning. A total of 48 English-speaking participants carried out a computer-based Portuguese pronunciation tutorial, followed by an operation span task (WM span task). They were then
tested on their recall of the Portuguese letter-sound rules that they had learned during the tutorial. It was found that participants with higher WMC showed higher accuracy in producing a new set of Portuguese words than those with lower WMC, and produced more Spanish-like errors than English-like errors. They interpreted both findings as indicating higher WMC learners’ superior ability to inhibit their L1 while learning their L2 Portuguese.

Contrary to the findings mentioned above, Juffs (2005; see also Juffs, 2004) found very low correlations between WMC measured via two versions of reading span task (learners’ L1 and L2) and the time it took ESL learners to read and process the embedded verb from the wh-subject extraction, with correlation coefficients ranging from -.15 to .08. The native control group was not an exception, showing a very low correlation coefficient between their WMC and the reading time ($r = -.20$). These results suggest that WMC may not relate to some aspects of syntactic processing (whether L1 or L2). Or it may be because WM assessed via reading span tasks may not reflect WM specifically required for syntactic processing, as Caplan et al. (2007) suspect that “one consideration regarding the finding that ‘standard’ WM capacity does not predict WM utilization in syntactic processing is that ‘standard’ WM measures are measures of WM capacity and the measures of on-line processing are measures of utilization” (p. 294).

In interaction research, Trofimovich et al. (2007) conducted a study to investigate whether and how individual differences in attention control, phonological memory, working memory, and analytic ability affect the extent to which learners notice and benefit from recasts. Grammatical (English possessive determiners) and lexical (English intransitive and transitive verbs) were selected as linguistic targets. Thirty-two adult Francophone ESL learners participated in a series of tasks including computer-based treatment/test activities, a non-word recognition task to measure phonological memory, the letter-number sequencing subtest from the
Wechsler Adult Intelligence Scale (WAIS-III) to measure WM, the Trail Making Test to measure attention control, and the grammatical sensitivity test (Part IV of the French version of the Modern Language Aptitude Test) to measure language analytic ability. Their results, albeit somewhat mixed, showed the benefits of recasts on the targets and some impacts of individual differences in attention control, analytical ability, and phonological memory on grammatical accuracy and/or lexical learning. However, the study found no significant correlations between noticing and any of the four factors operationalized. Furthermore, individual differences in WM were not found to have played any significant role in the learning of the targets.

Additionally, the role of PSTM in L2 learning has also been a topic of much research and related discussion (Baddeley, Papagno, & Vallar, 1988; Ellis, 1996; Ellis & Sinclair, 1996; Ellis & Schmidt, 1997; Kormos & Sáfár, 2008; Papagno, Valentine, & Baddeley, 1991; Service, 1992; Service & Kohonen, 1995; Williams & Lovatt, 2003). It has been evidenced that PSTM plays an important role in L2 vocabulary learning (Baddeley et al., 1988; Papagno et al., 1991; Service, 1992; Service & Kohonen, 1995) and in the process of acquiring chunks of language (Ellis, 1996). Furthermore, Ellis (1996) argued that individual differences in PSTM contribute to learners’ sequencing ability, which in the end causes individual differences in “learners’ facility to acquire second-language grammar” (p. 92). Ellis and Schmidt (1997) provided evidence for this argument by showing that phonological working memory capacity was closely related to the acquisition of morphosyntactic features. Similarly, Ellis and Sinclair (1996) observed that phonological rehearsal yielded a wide range of benefits in learning Welsh including vocabulary, pronunciation, and rule learning (e.g., phonological changes of the Welsh soft mutation). Additional support for the beneficial effects of PSTM on L2 learning comes from Williams and Lovatt’s (2003) study in which they found statistically significant correlations between PSTM
and ultimate rule learning ($r = .545$, $p < .05$ in Experiment 2 after the language background variable was partialled out), and between PSTM and memory for morpheme combinations ($r = .567$, $p < .01$), and between memory for morpheme combinations and ultimate rule learning ($rs = .468$ and $.641$ in Cycles 4 and 5, respectively), suggesting that PSTM influences learners’ rule learning ability both directly and indirectly.

In sum, as discussed above, WMC has emerged in cognitive psychology as a topic of crucial importance, engendering numerous studies on its relation to varied cognitive behaviors. To reiterate, what has been evidenced and confirmed in most of these studies is that WMC is strongly correlated with higher-order cognitive behaviors (e.g., L1 language comprehension, reasoning, general fluid intelligence, L2 learning, etc.). Regarding the present study, if WMC is related to noticing and modified output production, as indicated in Mackey et al. (2002, 2010), it can be reasonably speculated that the beneficial effects of corrective feedback may, to some extent, depend on individual differences in WMC, as supported by Sagarra (2007) in the case of recasts. One recent study (Goo, in press) investigated whether and how WMC mediates the extent to which learners benefit from different types of corrective feedback (recasts and metalinguistic feedback) in L2 learning, the summary of which is provided below in detail. In fact, the present research project was designed as an expanded version of Goo’s (in press) study based on his findings.

**Goo (in press)**

The study was designed to compare recasts with metalinguistic feedback in terms of their effects on L2 learning and, more importantly, to investigate whether and how WMC mediates the extent to which learners can benefit from recasts and/or metalinguistic feedback. The research questions addressed in the study are as follows.
Research Questions

RQ1: Is there any difference between recasts and metalinguistic feedback in their effectiveness at facilitating L2 learning?

RQ2: Is the efficacy of recasts and/or metalinguistic feedback on L2 learning mediated by individual differences in WMC?

Method

The study employed a pretest-treatment-posttest design. Fifty-four intermediate-level Korean EFL learners from 6 classes attending a university participated in this quasi-experimental study: recast (n = 14), metalinguistic feedback (n = 32), and control (n = 8). The English that-trace filter (e.g., *Who do you think that likes Mary?) was selected the linguistic target. The learners carried out two one-way information gap activities in a classroom setting. In each treatment session, they were presented with a drawing depicting an event (a graduation party for the first treatment session and an art gallery scene for the second treatment session) that was projected on a large screen in front of the class. All learners received a handout containing 15 sentences with a blank in each (10 critical items and 5 distractors) describing as many items (people or objects) shown in the drawing (e.g., John thinks that __________ is reading a newspaper.). Learners were instructed to ask a question about the name of each item, with which they needed to fill in the blank, to complete a given sentence. Then the NS teacher (also the researcher in the second session) provided relevant information. Learners in the treatment conditions received corrective feedback either in recasts (e.g., Who does John think is reading a newspaper?) or in metalinguistic feedback (e.g., Don’t use the conjunction “that” when you ask about the subject of the subordinate clause) depending upon their group affiliation whenever they produced nontargetlike sentences. For the metalinguistic group, the learners’ understanding
of such grammatical terms as *subject*, *subordinate clause*, and *conjunction* was checked before the first treatment session.

Two L1 WM span tasks were administered: reading span (RSPAN) and operation span (OSSPAN) tasks. Both span tasks were modified versions developed based on well-established previous span tasks, using MicroSoft PowerPoint. Both span measures were timed tasks, that is, each slide was programmed to automatically appear/disappear within a certain time frame determined in a pilot. The RSPAN included 42 Korean sentences that described various family relations and how family members and relatives should be called. All Korean consonants were used as to-be-recalled items and equally distributed (but not twice in the same set). The sentences and consonants were distributed in sets of 2, 3, 4, or 5 sentences. Set sizes were randomized. As for the OSPAN, a total of 42 mathematical operations (e.g., “(2 x 3) +1 =?”) were generated. As in the RSPAN, the Korean consonants were also utilized as to-be-remembered items and equally distributed (again not twice in the same set). Addition, subtraction, multiplication, and division were all counterbalanced. In both span tasks, when learners were presented with the recall cue (???) at the end of each set, they were required to recall the consonants shown in the preceding set and write them down in the order of presentation on the answer sheet.

Two types of dependent variable measures were used: timed grammaticality judgment tests (GJT) and written production tests. For each measure, two types were developed (Type A and Type B, one for the pretest and the other for the posttest). Included in each GJT were twenty-four pairs of declarative-interrogative sentences on as many slides (MS PowerPoint): 16 critical items and 8 distractors. Each pair consisted of one declarative sentence with an underlined noun phrase (NP) in it and one interrogative sentence purporting to ask about the underlined NP. Learners were later asked to correct the interrogative sentences that they judged ungrammatical.
In the written production test, learners were instructed to produce an interrogative sentence asking about the NP underlined in each sentence provided (the underlined NP was either the subject or the object of the embedded clause): 15 target sentences designed to elicit *wh*-movement involving the *that*-trace filter and 5 distractors designed to elicit *wh*-movement from the object position of the embedded clause.

**Results**

The Pearson correlation between the two WM span measures (on the recall part) showed statistical significance with a large effect size ($r = .580, p < .01$), indicating the two tasks measured a similar cognitive construct (i.e., WM). The alpha was set at .05 for all statistical tests. First, learner performance on the WM measures and the pretests was submitted to ANOVAs. No significant between-groups differences were observed in learner performance on the WM span tasks, $F (2, 51) = .467, p > .05$ ($p = .629$) for RSPAN, $F (2, 51) = .431, p > .05$ ($p = .652$) for OSPAN, and the pretests, $F (2, 51) = .808, p > .05$ ($p = .451$) for the GJT, $F (2, 51) = .335, p > .05$ ($p = .717$) for the written production test. That is, all three groups were more or less homogeneous in WM and their pre-experimental knowledge on the English *that*-trace filter. In order to further check whether WM should be included as a covariate in principal analyses, simple linear regression analyses were performed on each posttest as an outcome variable with each WMC measure as a predictor variable. None of the regression analyses showed statistical significance ($p > .05$), indicating WMC was not significantly related to the overall posttest performance of all three participating groups. Due to these nonsignificant results, mixed ANOVAs with Time (pretest-posttest) as a within-subjects variable and Group (recast vs. metalinguistic vs. control) as a between-subjects variable, rather than mixed ANCOVAs, were conducted as omnibus/principal analyses with respect to the first research question.
A two (Time: pretest-posttest) x three (Group: recasts, metalinguistic, control) mixed
ANOVA on the GJTs found a significant main effect of Time, $F(1, 51) = 20.193, p < .05$ ($p = .000, \eta^2_p = .284$), and of Group, $F(2, 51) = 4.651, p < .05$ ($p = .014, \eta^2_p = .154$), but no
significant Time x Group interaction effect, $F(2, 51) = 2.287, p > .05$ ($p = .112, \eta^2_p = .082$). All
three groups performed significantly better on the posttest than on the pretest. A significant
between-groups difference was obtained in the Brown-Forsythe procedure on the posttest, $F(2, 40.597) = 6.382, p < .05$ ($p = .004$). Dunnett’s T3 post hoc pairwise comparisons indicated that
both experimental groups significantly outperformed the control group, ($p = .014$ for the recast
vs. control groups and $p = .001$ for the metalinguistic vs. control groups), but no statistically
significant difference was observed between the two experimental groups ($p = .894$).

Another 2 x 3 mixed ANOVA conducted on the written production tests found a
significant main effect of Time, $F(1, 51) = 38.240, p < .05$ ($p = .000, \eta^2_p = .429$), of Group, $F(2, 51) = 5.651, p < .05$ ($p = .006, \eta^2_p = .181$), and of Time x Group interaction, $F(2, 51) = 5.061, p < .05$ ($p = .010, \eta^2_p = .166$). Both experimental groups, but not the control group, performed
significantly better on the posttest than on the pretest. A statistically significant between-groups
difference in their posttest performance was also found in the Brown-Forsythe procedure, $F(2, 29.299) = 7.412, p < .05$ ($p = .002$). According to Dunnett’s T3 post hoc pairwise comparisons,
both experimental treatment groups significantly outperformed the control group ($p = .003$ for
the recast vs. control groups, $p = .000$ for the metalinguistic vs. control groups), but again no
significant difference was obtained between the two experimental groups ($p = 1.000$). The Time-
by-Group interaction effect was attributable to the fact that the metalinguistic group performed
better than the recast group on the pretest, but the recast group outperformed the metalinguistic
group on the posttest.
To answer the second research question, simple linear regression analyses were performed with each WM span as a predictor variable and pre-to-post gain scores (normalized gain scores, see Hake, 1998) as an outcome variable. WMC measured via the RSPAN task was found to be significantly associated with the improved performance of the recast group on both dependent variable measures: $F(1, 12) = 10.884, p < .05, \beta = .690 (p = .006, R^2 = .476, \text{adjusted } R^2 = .432)$ and $F(1, 12) = 8.478, p < .05, \beta = .643 (p = .013, R^2 = .414, \text{adjusted } R^2 = .365)$ for the GJT and the written production test, respectively. Similarly, WMC measured via the OSPAN task also significantly predicted the improved performance of the recast group on both measures: $F(1, 12) = 5.668, p < .05, \beta = .566 (p = .035, R^2 = .321, \text{adjusted } R^2 = .264)$ for the GJT and $F(1, 12) = 5.933, p < .05, \beta = .575 (p = .031, R^2 = .331, \text{adjusted } R^2 = .275)$ for the written production test.

With regard to the metalinguistic group, WMC measured via the RSAPN task was not significantly related to the observed beneficial effects of metalinguistic feedback on learner performance on either dependent measure: $F(1, 30) = 2.004, p > .05, \beta = .250 (p = .167, R^2 = .063, \text{adjusted } R^2 = .031)$ for the GJT and $F(1, 30) = .075, p > .05, \beta = .050 (p = .786, R^2 = .008, \text{adjusted } R^2 = -.031)$ for the written production test. Likewise, the improved performance of the metalinguistic group was not found to be mediated by WMC measured via the OSPAN task: $F(1, 30) = 1.151, p > .05 \beta = .192 (p = .292, R^2 = .037, \text{adjusted } R^2 = .005)$ for the GJT and $F(1, 30) = .239, p > .05, \beta = -.089 (p = .629, R^2 = .008, \text{adjusted } R^2 = -.025)$ for the written production test.

In sum, the experimental groups significantly outperformed the control group on both dependent variable measures, but no significant difference was found between the two groups in either measure (Metalinguistic > Control, Recast > Control, Metalinguistic feedback = Recast).
WMC significantly positively predicted the pre-to-post development of the recast group, but not of the metalinguistic group.

**Discussion**

The findings suggest that recasts are as effective as metalinguistic feedback, and that WM may mediate the extent to which recasts, but not metalinguistic feedback, benefit L2 learning. The nonsignificant finding between the two groups, according to the researcher, may be attributable to the blocking of modified output opportunities. Considering recasts are not as effective in eliciting modified output as prompts such as metalinguistic feedback (e.g., Lyster & Ranta, 1997; Panova & Lyster, 2002), and given previous research that indicates the importance of modified/pushed output as a potentially contributing factor to L2 learning (e.g., McDonough, 2005; Nobuyoshi & Ellis, 1993; Swain, 1985, 1995, 2005), the blocking of modified output opportunities in the study may have been much more detrimental and damaging to the metalinguistic group than to the recast group. Also, the recast group benefitted from the blocking of modified output opportunities as recasts were provided in a context similar to a somewhat controlled experiment, although administered in a classroom setting, which may have furthered learner noticing by drawing more attention to the target syntactic regulation than in a much less controlled classroom setting with modified output opportunities allowed.

Regarding the findings in relation to the relationship between WMC and corrective feedback, the researcher suggested that the noticing of recasts on the violation of the *that*-trace filter, namely, the noticing of the deletion of *that* in full recasts, may have been attention-demanding and thus required executive attention processes, considered to be a crucial component of WMC (see, for instance, Engle, 2002; Engle, Kane, et al., 1999; Kane et al., 2007 for their executive attention theory). He further noted that the noticing of recasts during interaction
necessitates cognitive control of attentional resources because it requires learners to engage in
cognitive comparisons of their original non-targetlike utterances and targetlike utterances
provided through recasts and, of course, maintain the flow of communication at the same time.
This is particularly the case in the study, according to the researcher, considering some features
and characteristics of the type of recasts provided and of the target used in the experiment.
Recasts provided during interaction were long and full recasts more likely to go unnoticed as
compared to short/partial recasts. Also, because the conjunction that is a non-salient and non-
meaning-bearing constituent that does not affect comprehension, learner attention may not have
been directed at the presence or absence of that during meaningful interaction. In addition, given
the fact that the conjunction that can be deleted in various instances, it may have been quite
challenging for some learners (probably those with lower WMC less able to control attentional
resources compared to those with higher WMC) to notice the corrective nature of recasts. Goo
pointed out that these characteristics of the target and recasts used in the study likely maximized
the implicit dimension of recasts, in which individual differences in cognitive control are
inevitably involved, leading to the significant relationship between WMC and the effectiveness
of recasts in the acquisition of the syntactic phenomenon, the English that-trace filter.

Limitations and Future Research

The sample size of the recast group (n = 14) included in the data analyses is too small to
make any strong/generalizable claim. Both dependent variable measures employed in the study
(i.e., GJTts and written production tests) were designed to test learners’ grammatical knowledge
of the English that-trace filter. Other measures including oral performance tests need to be
utilized to provide more reliability for findings in future research. Also, because the type of
structure may differentially affect the effectiveness of corrective feedback (e.g., Ellis, 2007;
Iwashita, 2003; Leeman, 2003; Long et al., 1998; Sagarra, 2007), other structures in English or structures in other foreign/second languages, should be tested in future research. Also, other issues/topics to be considered in future research include whether other implicit forms of feedback with low noticeability are also related to WMC, whether L2 WM span tasks result in similar findings, and whether individual differences in WMC also predict or mediate the efficacy of short or partial recasts that have been found to be relatively easy to notice.

**WM, Attention Control, and Corrective Feedback**

Goo’s (in press) findings confirm the assumption that higher-span learners are able to actively maintain negative evidence (and targetlike forms) provided via recasts in a readily accessible state for future reference/use while focusing on meaning during interaction, due to their superior use of cognitive control of attentional resources (e.g., Conway et al., 2001; Engle, 2002; Engle, Kane et al., 1999; Engle, Tuholski et al., 1999; Kane & Engle, 2000, 2003; Kane et al., 2001, 2007; Rosen & Engle, 1997, 1998), whereas lower-span learners are more susceptible to interference (i.e., being distracted by meaning during interaction) due to their coarsely-grained attention-control mechanisms. Therefore, lower-span learners are less able to sustain negative evidence in an accessible form than their higher-span counterparts. Or lower-span learners are more likely to forget negative evidence on a specific linguistic feature during the time spent focusing on meaning than higher-span learners, as the task-switching account might suggest (Towse & Hitch, 1995, 2007; Towse et al., 1998). However, this does not seem to be the case as to metalinguistic feedback in learning the English *that*-trace filter. Also given the type of structure may play a role in L2 learning via recasts (e.g., Iwashita, 2003; Leeman, 2003; Long et al., 1998), it appears to be necessary to use different types of target structures to confirm or disconfirm Goo’s findings, as he suggested.
Another interesting issue is whether individual differences in cognitive control reflected in performance on simple attention control tasks that put no demands on memory capacities are correlated with performance predicted by WMC. As mentioned earlier, high-WMC individuals are better able to control their attention and maintain task goals than those with low WMC in such simple attention-control tasks as dichotic listening tasks (Conway et al., 2001), Stroop tasks (Kane & Engle, 2003) and antisaccade tasks (Kane et al, 2001). For instance, Conway et al. (2001) found 65% of low-span participants, according to the results of their retrospective questionnaire, reported hearing their name in the irrelevant message (distractions) provided through their left ear while repeating words in the goal-relevant message presented to their right ear, whereas only 20% of high-span participants reported hearing their name, clearly showing their high-span participants used much more controlled attention in maintaining and implementing the task than low-span participants. Similarly, Kane et al. (2001) found that their high-span participants were faster and more accurate than low-span participants in the antisaccade task, and that high-span participants made fewer eye-movement errors and were able to recover from eye-movement errors more quickly than low-span participants, providing further evidence of high-span individuals’ superior use of controlled attention. Additional evidence was also obtained in Kane and Engle’s (2003) study in which they showed that low-span subjects were more susceptible to Stroop interference effects than high-span subjects, committing more color-naming errors than did their high-span counterparts, especially in the 75% or 80% congruent conditions. If in fact this type of simple attention control is at the epicenter of WMC variation, as argued by Engle and his colleagues (e.g., more recently in Kane et al., 2007), then performance on those simple attention control tasks (e.g., the Stroop task) should in essence be
correlated with (or predict) performance that is predicted by WMC. This assumption was also tested as a part of the present research project.
Chapter III: The Present Experiment

Research Questions

The present dissertation research as an expanded version of Goo’s (in press) study is designed to explore the relationship between the type of corrective feedback and the type of structure in interaction-driven L2 learning (two different target structures in terms of their linguistic characteristics and communicative values and two types of corrective feedback). The study is also intended to examine whether and how individual differences in cognitive capacities mediate the efficacy of corrective feedback on learner performance. Below are two research questions addressed in the present research.

RQ1: What are the relationships between the type of corrective feedback and the type of target structure in interaction-driven L2 learning?

RQ2: How is the effectiveness of corrective feedback mediated by individual differences in cognitive capacities?

Method

Participants

A total of 83 Korean EFL learners participated in the present study (male = 27, female = 56). Considering learners from the same population are likely to have a similar range of cognitive capacities, for the present study, as in Kane et al. (2001) and Kane et al. (2004), learners were recruited from 3 different universities (\( n = 68 \)) and 2 elementary schools (\( n = 15 \), teachers). The university learners were English education, English literature, special education, or computer science majors, most of whom were taking an English education course at the time of the study. Some of them were taking English conversation classes at private English language institutes. They were deemed as intermediate learners of English based either on their professors’
comments or their own comments on their overall proficiency including conversation skills. All 15 elementary school teachers either had taken or were taking teacher training courses specifically designed for elementary English teachers and were teaching and studying English in their own schools at the time of the study. On their own judgment, their proficiency was also considered intermediate. All participants were randomly assigned to one of three groups/conditions in this experimental study, Group 1 (n = 24), Group 2 (n = 27), and Group 3 (n = 32); Groups 1 and 2 are treatment conditions, and Group 3 is a control condition. Their ages ranged from 19 to 40 (M = 24.29, SD = 4.34). They began learning/studying English at age 11.75 on the average (SD = 2.53) and studied for 9.51 years on the average (SD = 2.10).

Table 2

<table>
<thead>
<tr>
<th>Gender</th>
<th>Age</th>
<th>Onset</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
</tr>
<tr>
<td>Male</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group 1 (n = 24)</td>
<td>7</td>
<td>17</td>
<td>23.96</td>
</tr>
<tr>
<td></td>
<td>11.38</td>
<td>2.26</td>
<td>10.48</td>
</tr>
<tr>
<td>Group 2 (n = 27)</td>
<td>10</td>
<td>17</td>
<td>24.89</td>
</tr>
<tr>
<td></td>
<td>11.84</td>
<td>2.93</td>
<td>9.16</td>
</tr>
<tr>
<td>Group 3 (n = 32)</td>
<td>10</td>
<td>22</td>
<td>24.03</td>
</tr>
<tr>
<td></td>
<td>11.96</td>
<td>2.28</td>
<td>9.02</td>
</tr>
<tr>
<td>Total (N = 83)</td>
<td>27</td>
<td>56</td>
<td>24.29</td>
</tr>
<tr>
<td></td>
<td>11.75</td>
<td>2.53</td>
<td>9.51</td>
</tr>
</tbody>
</table>

*Onset means the age when they first began to learn English either at school or at home.*

*Length means the number of years for which they had studies.*

Of those 83 learners in the present study, 21 learners (25.30%) had been to English-speaking countries such as the U.S. (n = 5), U.K. (n = 4), Canada (n = 3), Australia (n = 8), and New Zealand (n = 1) mostly for traveling and/or short-term language courses (one or six months at the longest). Also, sixty learners (72.29%) had studied other foreign languages such as Chinese (n = 18), Japanese (n = 34), French (n = 2), German (n = 2), and Spanish (n = 4) during
their high school years. Each participant received a gift certificate worth 10,000 won, which is equivalent to approximately 9 U.S. dollars.

**Linguistic Targets**

**Targets.** Two linguistic targets selected for the present study are the English *that*-trace filter (see Example 1) and the past unreal/counterfactual conditional (see Example 2). They were chosen for a comparison because they are somewhat distinct in terms of their underlying linguistic features and characteristics, which is of importance given the present study is specifically designed to explore the impact of the type of target structure on beneficial impacts of corrective feedback on L2 learning. Interpreting results by comparing and contrasting these two distinct linguistic phenomena likely provides a clearer picture of the role that the type of target may play than when two similar structures are compared/contrasted in terms of their role. The English *that*-trace filter involves a *wh*-movement (purely syntactic), whereas the past unreal/counterfactual conditional requires morphosyntactic knowledge and semantic understanding.

Example 1

a. *Who do you think that likes Mary?*

   Who do you think likes Mary?

b. *Who do you think that won the game?*

   Who do you think won the game?

Example 2

a. If I had had enough time yesterday, I would have done my homework.

b. If John had not missed his bus this morning, he would not have been late for class.
The understanding of the English that-trace filter is accompanied by knowledge about wh-movement and a relevant subject-object asymmetry (see Juffs & Harrington, 1995 and Schachter & Yip, 1990 for their experiments/discussions on the subject-object asymmetry in terms of processing differences). The Empty Category Principle (ECP) provides explanations on this issue of the that-trace filter or subject-object asymmetry in wh-movement. According to the ECP, a trace must be either theta-governed (governed and theta-marked by a head) or antecedent-governed (governed by a co-indexed XP) (Chomsky, 1986, as cited in Haegeman, 1994). When wh-movement originates from the subject position preceded by the conjunction (or complementizer) that, unless it is deleted, the conjunction that blocks a trace from being antecedent-governed. In Example 3a, the conjunction that blocks the intermediate trace (t’i) in [Spec, CP2] from antecedent-governing the trace (ti) in the subject position, a violation of the ECP, which makes the sentence ungrammatical (compare Examples 3a and 3b).

Example 3
a. *[CP1 Who do [IP1 you think [CP2 t’i that [IP2 t i likes Mary]]]]

b. [CP1 Who do [IP1 you believe [CP2 t’i [IP2 t i likes Mary]]]]

If wh-movement is from the object position, however, the use/deletion of that does not make any difference in terms of grammaticality because the object trace (ti) is theta-governed by [V, VP], that is, there is no need for the object trace (ti) to be antecedent-governed by the intermediate trace (t’i), which reflects a subject-object asymmetry as to the use/deletion of the conjunction (or complementizer) that in wh-movement, as shown in Example 4.

Example 4
a. [CP1 Who do [IP1 you believe [CP2 t’i that [IP2 Sarah loves t i ]]]]

b. [CP1 Who do [IP1 you believe [CP2 t’i [IP2 Sarah loves t i ]]]]
English conditionals have been used as a target of acquisition in quite a few studies (e.g., McDonough, 2004; Rosa & O’Neill, 1999). Converting two related clauses into a conditional sentence usually involves several changes such as a semantic change (from negative to positive or vice versa), a changes to the tense or aspect of the if clause, a change to the tense and/or aspect of the main or result clause, and the inclusion of would or other modal verbs in the main or result clause. Especially, generating a past unreal/counterfactual conditional (see, for different terms on the structure, Celce-Murcia & Larsen-Freeman, 1999; Cowan, 2008; Swan, 2005) involves most of the above-mentioned linguistic changes, which gives rise to a longer sentence compared to the original clauses due mainly to their changes from simple past to past perfect in the if clause and from simple past to would have + past participle in the main or result clause (see Example 5). Therefore, converting two clauses into a past unreal conditional sentence entails much mental processing in preparation and even while producing it (especially in oral production), which makes it quite challenging to fully acquire the structure.

Example 5

a. I missed the bus, so I called a cab this morning.

   If I had not missed the bus, I would not have called a cab this morning.

b. I had a headache this morning, so I didn’t go to school.

   If I hadn’t had a headache this morning, I would have gone to school.

Rationale. The English that-trace filter was selected as one of the two linguistic targets based on three reasons: communicative redundancy, susceptibility to individual differences in WMC, and rare exposure to the target. First, it is communicatively redundant because it is purely syntactic and does not affect comprehension of a given wh-question although it is violated. This communicative redundancy of the target necessitates focus on form intervention through
corrective feedback. Second, in addition to its communicative redundancy, this particular structure requires selective attention that helps notice the deletion of the non-meaning-bearing element “that” only when the subject (not the object) of the embedded clause is moved, which makes the structure difficult to acquire unless attentional resources are properly controlled. Considering cognitive control is a critical component of WM (Engle, 2002; Engle, Kane et al., 1999; Engle, Tuholski et al., 1999; Kane et al., 2007), it can be assumed that individual differences in WMC likely affect the acquisition of this particular target structure during interaction. Third, this syntactic phenomenon is neither specifically dealt with in ESL/FL grammar books nor explicitly taught in classroom contexts. In addition, as English wh-questions involving wh-movement across two clausal boundaries from the subject position of the embedded clause are not frequently used in everyday conversation, L2 learners’ exposure to the that-trace filter in naturalistic/incidental learning contexts is also limited. Thus, it appears crucial to provide L2 learners with some opportunities to receive focus on form intervention in order to acquire the target construction. Also, for the same reason, the selection of this particular syntactic phenomenon was assumed, to some extent, to minimize impacts of potentially confounding factors (e.g., prior learning or exposure).

As for the second target, all the changes necessary to produce a past unreal conditional sentence are critical to deliver intended meaning to interlocutors and maintain the flow of communication as a whole. Thus, it is of crucial importance to fully understand the relevant rule and be able to convert two clauses into a past unreal conditional sentence. However, because of this very nature of making multiple changes to the original clauses, L2 learners are burdened with a heavy cognitive load, especially when they produce a past unreal conditional sentence in oral communication. In other words, producing a past unreal conditional sentence does require
L2 learners to engage in careful preparation and intense and rapid processing in order to supply a grammatical utterance without sacrificing their intended meaning. This is where individual differences in cognitive capacities may play a role. L2 learners must process all necessary changes rapidly and correctly by paying careful attention to those linguistic areas required to form an appropriate sentence, which appears to be somewhat similar to multitasking. The success of this cognitive operation, therefore, may depend on how accurately and selectively L2 learners manage their attentional resources while processing the needed changes. This processing aspect makes the structure suitable for the present investigation given WMC (and obviously attention control) allegedly involves cognitive control of attentional resources. Regarding L2 learners’ exposure to the target, although it is explicitly dealt with in ESL/FL grammar books, its use is relatively infrequent. Especially, for learners in an EFL context, their exposure (also out-of-experiment exposure, for that matter) is limited, and opportunities to use it are quite rare, due to which L2 learners still fail to acquire the structure. Thus, focus on form intervention by providing opportunities for learners to produce past unreal conditional sentences and receive corrective feedback on their nontargetlike utterances seems necessary to help learners fully acquire the target rule/structure.
The present study examined how the above-mentioned similarities and differences (see Table 3) are related to the effectiveness of different types of corrective feedback, especially how their differences in linguistic characteristics and communicative value contribute to the relative efficacy of recasts and metalinguistic feedback on L2 learning.

**Experimental Design**

The present study employed a pretest-treatment-posttest-delayed-posttest design. During the treatment, corrective feedback was provided for the learners in the two experimental conditions (Groups 1 and 2) on their erroneous utterances with regard to each of the two targets. As seen in Table 4, Group 1 received recasts on their nontargetlike output that violated the *that*-trace filter and metalinguistic feedback on the past unreal conditional, whereas Group 2 received metalinguistic feedback on their nontargetlike utterances involving the *that*-trace filter and recasts as to the past unreal conditional. No feedback was provided for Group 3 (control group).
Table 4

Experimental Treatment Design

<table>
<thead>
<tr>
<th>Group 1 (n = 24)</th>
<th>English that-trace filter</th>
<th>English Past unreal conditional</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group 2 (n = 27)</td>
<td>Metalinguistic feedback</td>
<td>Recasts</td>
</tr>
<tr>
<td>Group 3 (n = 32)</td>
<td>No feedback</td>
<td>No feedback</td>
</tr>
</tbody>
</table>

*Note. As illustrated in the section on the treatment activity later in this chapter, the two target structures were incorporated into one treatment activity in such a way that the learners needed to alternate their focus on the two targets as they were required to produce output involving one target immediately followed by output involving the other target. That is, potential order effects (the order of treatments/structures) were internally controlled for in this single-task experimental treatment activity.*

To further elaborate, as regards the English that-trace filter, a full recast operationalized as a reformulated sentence with a change or two made to the learner’s nontargetlike utterance that violated the that-trace filter was provided for the learners in the recast condition (see Example 6). Metalinguistic feedback operationalized as linguistic information (containing grammatical terms) related to the ill-formedness of the learner’s original utterance was provided for the metalinguistic condition whenever the learner’s utterance violated the that-trace filter (see Example 7). For the learners of the metalinguistic group, their understanding of such grammatical terms as *subject, conjunction*, and *subordinate/embedded clause* was checked and confirmed before the treatment. If needed, the terms were explained in or translated into L1 (Korean) for those, among the learners of the metalinguistic group, who did not show a full understanding of them.

Example 6

Learner: Who does Sang-goo think that broke into Mary’s house last night?

Researcher: Who does Sang-goo think broke into Mary’s house last night? It’s Mike, because
Mike said to Sang-goo…

Example 7

Learner: Who does Joon-ho think that stole them from the jeweler’s last week?

Researcher: You are not supposed to use the conjunction that when you ask about the subject of the subordinate clause or embedded clause. It’s Alex, by the way, because Alex said to Joon-ho…

With regard to the past unreal conditional, the recast group received either a full recast or a clause-level partial recast depending upon the location of the learner’s nontargetlike output (i.e., if clause, main/result clause, or both clauses). For instance, in Example 8 only the if clause was reformulated with one change (from had have to had had), and its reformulated version is provided, whereas a recast is provided on the learner’s output in the main clause in Example 9.

When both clauses were ungrammatical, a recast on the entire sentence was provided (see Example 10).

Example 8

Learner: Who does Young-soo think has cancer?

Researcher: It’s Tom because Tom said to Young-soo…

Learner: If I had have a physical check-up last month, I would have received medical treatment earlier.

Researcher: If I had had a physical check-up last month. Next, Dong-hyun.

Example 9

Learner: Who does Joon-ho think that stole them from the jeweler’s last week?

Researcher: You are not supposed to use the conjunction that when you ask about the subject of the subordinate clause or imbedded clause. It’s Alex, because Alex said to Joon-ho,
Learner: If I had had a gun the other day, I would steal that expensive watch from the jeweler’s.

Researcher: I would have stolen that expensive watch from the jeweler’s. Next, Kwang-mo.

Example 10

Learner: Who does Young-soo think stole them from the jeweler’s last week?

Researcher: It’s Mike, because Mike said to Young-soo,

Learner: If my girlfriend didn’t want it, I would not steal it from a jeweler’s.

Researcher: If my girlfriend had not wanted it, I would not have stolen it from a jeweler’s. Next, Joon-ho.

As for the metalinguistic condition, such relevant grammatical terms as past unreal conditional, if clause, main/result clause, past perfect, and past participle were checked with learners and explained to those who did not show a clear understanding of the terms before the treatment activity. The term past unreal conditional was translated into their L1, and only its communicative function was expounded, but not the actual morphosyntactic rule involved in the formation of a past unreal conditional sentence. As in the case of the recast condition, the amount of grammatical/metalinguistic information on the past unreal conditional provided for the metalinguistic condition was determined by the location of nontargetlike output (i.e., if clause, main clause, or both); see Examples 11, 12, and 13.

Example 11

Learner: Who does Joon-ho think stole them from the jeweler’s last week?

Researcher: It’s Alex, because Alex said to Joon-ho,

Learner: If I would had had a gun the other day, I would have stolen that expensive watch from the jeweler’s.
Researcher: You need the past perfect in the ‘if’ clause. Next Kwang-mo.

Example 12
Learner: If my ex-girlfriend had not bought it for me, I would not have had a chance to wear such an expensive watch.

Researcher: You need ‘would’ plus ‘have’ plus the past participle of the verb in the main clause.

Example 13
Learner: Who does Sang-goo think stole them from the jeweler’s last week?

Researcher: It’s John, because John said to Sang-goo,

Learner: If I met Sister Sara the other day, I would not think about stealing that expensive watch.

Researcher: This is a past unreal or past counterfactual conditional. You need the past perfect in the ‘if’ clause, and you need ‘would’ plus ‘have’ plus past participle of the verb in the main clause. Next, Dong-hyun.

In order to control for modified output opportunities, as in Goo (in press), Leeman (2003), and Long et al. (1998), because they may impact L2 learning (e.g., McDonough, 2005; Nobuyoshi & Ellis, 1993; Swain, 1985, 1995, 2005), opportunities for learners of both treatment conditions to modify their initial ungrammatical utterances were not blocked by the researcher’s prompts such as “Next, Kwang-mo” and “It’s Alex, because Alex said to Joon-ho…” immediately following the provision of corrective feedback (see the Examples above). The control group also participated in the interactional task (without feedback) and all three test sessions, but not the two cognitive tasks (i.e., the OSPAN task and the task-set switching test).
Tasks and Materials

Treatment task (Rashomon). Learners participated in an interactional task in which they were required to produce the two target forms. The basic structure of the task was inspired by a Japanese crime movie, *Rashomon*, directed by Akira Kurosawa in 1950. This murder mystery provides a unique storyline around a murdered samurai, for which it has been critically acclaimed for generations. Four witnesses give four different accounts of the murder and the motive for it, similarly structured but still different stories. As in the movie, each portion of the treatment task unfolded around an incident/event that happened in the past. Six people pointed at six different individuals as a main character involved in a given incident. This one-time treatment task consisted of three incidents/events of that kind: stealing expensive watches from a jeweler’s shop, breaking into Mary’s house, and having cancer. In the first scenario, to elaborate, six people had different ideas about who had stolen watches from a jeweler’s shop the previous week based on their own encounter with the person that they suspected had committed the crime. For instance, according to Cheol-soo’s account, Tom stole watches from the jeweler’s last week based upon what Tom had said to him at some point earlier in time, but Young-soo suspected Mike of the crime based on what Mike had said earlier, and the like.

The task took place in a dyad between each learner and the researcher. At the beginning of the actual task, the learner was presented with a letter-size drawing that depicted an incident/event and briefed by the researcher on it. The learner then received a handout from the researcher that included 6 sentences for each of seven incidents/events: two incidents for the oral pretest, three for the treatment, and the remaining two for the immediate posttest. Each sentence contained a blank in the subject position of the embedded clause, which needed to be filled in with a name (see Example 14).
Example 14 (Stealing expensive watches from a jeweler’s shop last week)

Cheol-soo thinks that ____________ stole them from the jeweler’s last week.

Young-soo thinks that ____________ stole them from the jeweler’s last week.

Joon-ho thinks that ____________ stole them from the jeweler’s last week.

Kwang-mo thinks that ____________ stole them from the jeweler’s last week.

Sang-goo thinks that ____________ stole them from the jeweler’s last week.

Dong-hyun thinks that ____________ stole them from the jeweler’s last week.

The learner was instructed to ask the researcher a question about the name of the person suspected of the incident, according to one of those six individuals who had their own ideas regarding who did it, necessary to complete the corresponding sentence. The learner was required to use the verbs in both the main clause and the subordinate/embedded clause (e.g., “think” and “steal” in Example 14). Then the researcher who already had relevant information provided the learner with the name of the suspect; for instance, according to Young-soo (in the scenario), Mike stole watches from the jeweler’s last week, thus the researcher gave the name “Mike” to the learner. Immediately after supplying the name of the presumed suspect (e.g., Mike in Young-soo’s opinion), the researcher showed the learner a drawing that described what the presumed suspect had said earlier to express his regret, discontent, or wishful thinking about a past event, of course, using a past unreal conditional sentence. The suspect’s words were provided in written form at the bottom of each drawing that depicted them, with two underlined blanks in the suspect’s past unreal conditional sentence (see Example 15): one in the if clause and the other in the main/result clause. Adjacent to each blank was a verb in a base form (and sometimes with “not”) provided in parentheses. The learner was instructed to produce a past unreal conditional sentence using the verbs. In order to lessen the learner’s cognitive burden, the
learner was further instructed to include “not” in his/her past unreal conditional sentences only when it was provided in parentheses, as in the case of Example 15.

Example 15

According to Young-soo,

Mike called me yesterday and said “Oh my god, I just stole a very expensive watch, the one that my girlfriend saw last Tuesday. If my girlfriend ___________ (not, want) it, I _______ (not, steal) it from a jeweler’s.”

Then the learner moved on to the next person and asked the researcher another question by producing an interrogative sentence involving the that-trace filter (e.g., *Who does Joon-ho think that stole them from the jeweler’s last week?). Again, after the learner identified the suspect, the researcher showed another drawing that described the suspect’s regret, discontent, and wishful thinking with his words in written form at the bottom of the drawing. The learner went through the same procedure for all those six accounts by six different individuals, producing six sentences involving the English that-trace filter and six past unreal conditional sentences for each incident/event. As mentioned above, three incidents/events were utilized for this intensive one-time interactional treatment. That is, each learner was provided with 18 opportunities to generate output containing each of the two linguistic targets, and of course, as many possible opportunities to receive corrective feedback (see Appendix A).

**Operation span task (OSPAN).** An operation span task was selected to measure WMC. Despite previous research findings that show high correlations between L1 and L2 versions of WM span measures (Mackey et al., 2002; Miyake & Friedman, 1998; Osaka & Osaka, 1992; Osaka et al., 1993; van den Noort et al., 2006), the selection of an operation span task was, as a precautionary measure, intended to prevent L2 knowledge and proficiency from influencing their...
span performance. For the same reason, the span task developed for the present experiment was specifically designed in such a way that little knowledge on L2 was required, as described below.

An operation span task is another complex span measure that has been widely utilized and accepted as tapping verbal WMC since Turner and Engle’s (1989) first version (e.g., Bunting, Conway, & Heitz, 2004; Conway et al., 2002, 2003; Goo, 2010, in press; Kane et al., 2004; Rosen & Engle, 1997, 1998; Tokowicz et al., 2004; Unsworth, 2007; Unsworth & Engle, 2007; also see Unsworth, Heitz, Schrock, & Engle, 2005 for a discussion of an automated version of the operation span task). As in other verbal WM span measures such as reading span and counting span tasks, it necessitates the operation of WM because it involves maintaining to-be-remembered items for later recall while solving math problems. For the present study, a timed operation span task modified from Unsworth et al.’s (2005) task, which showed good internal consistency reliability (alpha = .78) and test-retest reliability (.83), was developed, using Microsoft PowerPoint. A total of 75 simple mathematical operations (e.g., “(2 x 2) +1 = ?”) were generated. Following Kane et al. (2004), a set of English alphabetical letters (F, H, J, K, L, N, P, Q, R, S, T, Y) that are phonologically distinct were used as recall items and equally distributed, but not twice in the same set. The mathematical operations were distributed in sets of 3, 4, 5, 6, or 7 (three sets for each set size). Set sizes were randomized. Addition, subtraction, multiplication, and division were all counterbalanced.

A baseline subset of each set size consisted of three slides: one showing a mathematical operation, another showing an Arabic numeral (as a possible answer to the operation) and “Yes or No?” (presented in L1) intended to request learners to judge the correctness of the number as an answer to the operation, and the other showing an alphabetical letter to be remembered. In the actual task, learners silently read each mathematical operation that remained on a slide for 3
seconds, automatically followed by another slide that showed a supposed answer to the operation and a dichotic option “Yes or No?” Three seconds were provided for learners to decide whether the number provided was a correct answer to the operation and put a check mark on the answer sheet. On the following slide, an alphabetical letter, one of the selected letters mentioned above, appeared and stayed for 1 second, automatically followed by another mathematical operation or a recall cue (“????”). When learners were presented with the recall cue at the end of each set, they were supposed to recall the letters shown in the preceding set and write them down in the order of presentation on the other side of the answer sheet. For the recall portion of the operation span task, they were given 5 seconds for set size 3, 6.5 seconds for set size 4, 8 seconds for set size 5, 9.5 seconds for set size 6, and 11 seconds for set size 7, as determined in a pilot test. Three practice sets of size 2 were used before the actual task (see Appendix B for the math answer and recall sheets).

Figure 1. Sample presentation of each item in the OSPAN task.
**Attention control task.** Learners’ attention control was measured using the so-called task-set switching paradigm (e.g., Allport, Syles, & Hsieh, 1994; Rogers & Monsell, 1995; see Kiesel et al., 2010; Koch, Gade, Schuch, & Philipp, 2010; Logan, 2003; Monsell, 2003 for reviews on research on cognitive control in this paradigm). The term *task set* denotes “the organization of cognitive processes and mental representations that enable the person to act in accordance to task requirements” (Kiesel et al., 2010, p. 850). Therefore, it involves “a representation of a task goal (e.g., attend to stimulus color), a set of task-relevant stimuli (e.g., red and blue), a set of possible responses (e.g., press a left or a right key), and a mapping of stimuli-or stimulus categories-to responses” (Koch et al., 2010, p. 2). In a typical task switching experiment, participants are instructed to switch between two tasks. Cognitive/executive control in this paradigm is measured in terms of *switch costs* (Rogers & Monsell, 1995; see Altmann, 2007; Wylie & Allport, 2000 for more discussion on switch costs) represented by differences in reaction time (RT) and error rates between switch trials (trials involving a different task from the one that participants have just engaged in) and nonswitch trials (trials involving the same task that they have just executed). Among several task switching research techniques, the present study employed Rogers and Monsell’s (1995) alternating-runs paradigm, one of the most widely used techniques (e.g., Koch, 2005; Milán, Sanabria, Tornay, & González, 2005; Monsell, Sumner, & Waters, 2003). In the alternating-runs paradigm, tasks alternate every second trial (AABBAABB sequences); that is, participants either repeat the same task or switch from one task to another. Following this switching paradigm, in the present study, each learner was instructed to switch between a letter task (a consonant or a vowel) and a number task (an even number or an odd number) in the AABBAABB sequence.
The task-set switching paradigm developed for the present experiment consisted of four blocks of 64 trials, that is, a total of 256 trials were created with 192 actual trials and 64 practice trials. In each trial, the learner was presented with either a letter-number or number-letter pair (e.g., 9A, K6, U2, 3G, etc.). In the letter task, the learner was supposed to indicate whether a given letter shown in a letter-number or number-letter pair was a consonant or a vowel, disregarding the number shown together in the pair, by pressing “Z” on the computer if it was a consonant or “M” if it was a vowel. In the number task, the learner needed to indicate whether a given number in the pair was odd or even, disregarding the letter presented together in the pair, by pressing “Z” on the keyboard if it was an even number or “M” if it was an odd number. Thus, the learner needed to (and was advised to, for that matter) focus either on the letter or on the number in each pair, not both of them, depending upon a task cue presented before each stimulus pair (either a hash (#) indicating the letter task or a percent mark (%) the number task). The directions for the task were clearly explained in their L1 (Korean) before the practice block. All stimuli and cues were presented in black boldface, Times New Roman, 40 points in font size, on the center of the white screen on a Macintosh laptop computer, using SuperLab Pro 4.

In the actual task, the learner was first presented with a hash (#) signaling the letter task, which remained on the screen for 1 second; the first pair in each block was always for the letter, so each block began with a hash. A letter-number or number-letter pair then appeared and stayed either until the learner pressed “Z” or “M” or until 3 seconds passed by (a maximum amount of RT given to the learner). The learner was then led automatically to the next trial that began again with a hash indicating another letter task to be performed. After this second letter task, the learner engaged in two number tasks in a row, each of which began with a percent mark, followed by two letter tasks, and so forth. As such, the tasks alternated every two trials (e.g., LL-
NN-LL-NN…), and the learner went through the same process to finish all 64 trials included in each of 4 blocks (the first block was a practice block). The learner was instructed to respond to stimuli as quickly and accurately as possible. The learner took a short break after each block.

*Figure 2*. Sample presentation of AABB sequence in task switching.

**Assessment tasks.** Two dependent variable measures were developed and utilized to assess learner knowledge and performance on each of the two linguistic targets: oral production and grammaticality judgment tests. These two types of measures were selected as they have been assumed to provide potentially different aspects (and/or sources) of learner performance (see Ellis, 2004, 2005, 2006, 2009; Ellis et al., 2009 for discussions on implicit and explicit knowledge and relevant measures). Oral production tests during interaction supposedly measure a spontaneous
use of a select target item, whereas grammaticality judgment tests measure knowledge about the structures.

For each type of measure, three versions (A, B, and C) were created: one for the pretest, another for the posttest, and the other for the delayed posttest. The three versions were ordered in A-B-C, C-A-B, or B-C-A for the three test sessions. Approximately, one third of the learners followed the order of A-B-C, another third the order of C-A-B, and the remaining third the order of B-C-A. This was intended to minimize any nuisance factor related to the ordering of test versions. Each dependent variable measure is now described below.

**Oral production test.** As indicated above in the section about the treatment task, Rashomon, the oral production test was identical to the treatment task in every aspect except that obviously, no feedback was provided to the experimental groups, and that different incidents/events were introduced, and thus different sentences/drawings were utilized. Two incidents/events in version A were “Bank robbery” and “Going to the movies with Sue.” Included in version B were “Drinking Jay’s coke containing a cigarette butt” and “Smoking in a movie theater.” Finally, “Smashing a laptop computer” and “Murdering a pizza delivery guy” were used in version C (see Appendix A mentioned above).

**Grammaticality judgment test (GJT).** With regard to the English that-trace filter, 72 pairs of declarative-interrogative sentences were developed for all three versions of the test using Microsoft PowerPoint: 48 critical items and 24 distractors. Each pair consisted of one declarative sentence with an underlined noun phrase (NP) (either the subject or the object of the embedded/subordinate clause) and one interrogative sentence that asked about the underlined NP (see Example 15). The pairs/items were randomly distributed (stratified randomization) and grouped into three batches of 24 pairs (Types A, B, and C): one for the pretest, another for the
posttest, and the other for the delayed posttest. Of those 24 in each version, 16 critical pairs/items containing 8 grammatical and 8 ungrammatical interrogatives involved the that-trace filter, and the remaining 8 items (all grammatical) were distractors involving wh-movement from the object position of the embedded clause. In the actual test, learners were instructed to judge the grammaticality of the interrogative that asked about the NP underlined in the declarative sentence presented on each slide and to then put a check mark in the appropriate box within 10 seconds (grammatical, ungrammatical, or I don’t know), as determined in a pilot. For instance, Example 16a includes an ungrammatical interrogative that violates the that-trace filter, but the interrogative sentence in Example 16b conforms to it. Each new slide was programmed to appear every 10 seconds. The GJT slides were saved to the researcher’s computer prior to the actual tests and administered individually. Immediately after all the judgment pairs/items were presented, learners were instructed to return to the GJT slides and correct the interrogatives that they had judged ungrammatical (see Appendix C).

Example 16

a. They think that Jeff loved Jennifer.

*Who do they think that loved Jennifer?

b. They think that John is going to marry Sarah next month.

Who do they think is going to marry Sarah next month?

A similarly designed GJT was used for the other target structure, the past unreal conditional. A total of 48 pairs of declarative-conditional sentences were developed for all three versions used in the pretest, posttest, and delayed posttest. Of those 48 pairs, 16 pairs were presented in as many slides in each test session. Each pair consisted of two declarative sentences describing two related events that occurred in the past and one conditional sentence supposed to
be contrary to what was represented by the two declarative sentences. All verbs/modal auxiliaries used in the declarative sentences were formed in the past tense, and as such, the declarative sentences were supposed to describe past events that had actually happened; learners were clearly informed of this assumption before carrying out the conditional part of the pretest and again later reminded at the beginning of each posttest. One caveat specifically mentioned to the learners was that the modal verb could was used/included in some declarative sentences just as the past tense form of can meaning was/were able to only. Also, they were informed that two (or more) modal verbs were possible in some result clauses, and thus their judgment should not be based on which specific modal verb was used in the result/main clause (see Appendix D).

Of those 16 conditional sentences used in the GJT, half were grammatical, and the other half sentences were ungrammatical. Among the ungrammatical ones, four sentences were ungrammatical in both clauses (the if clause and the main/result clause), two sentences were problematic only in the if clause (simple past instead of past perfect), and the remaining two sentences were ungrammatical only in the main/result clause (would + main verb instead of would have + past participle); see Example 17.

Example 17
a. Declarative-conditional pair with both conditional clauses grammatical

I lost my passport. I asked for a new one.

If I had not lost my passport, I wouldn’t have asked for a new one.

b. Declarative-conditional pair with both conditional clauses ungrammatical

He broke his ankle last night. He did not work out this morning.

If he did not break his ankle last night, he would work out this morning.

c. Declarative-conditional pair with ungrammatical if clause
Jane had a car accident. She could not play soccer.

If Jane didn’t have a car accident, she could have played soccer.

d. Declarative-conditional pair with ungrammatical main/result clause

The director made that movie. My friends were able to recognize him.

If the director had not made that movie, my friends would not be able to recognize him.

In the actual test, as in the GJT on the *that*-trace filter, learners were instructed to judge the grammaticality of the conditional sentence in the pair presented on each slide and to put a check mark in the appropriate box within 15 seconds (*grammatical*, *ungrammatical*, or *I don’t know*). That is, the declarative-conditional pair was programmed to remain on the slide for 15 seconds, then a new slide appeared automatically. Again, all the GJT slides were saved to the researcher’s computer prior to the actual tests and administered individually. Immediately after all pairs were presented, learners were instructed to return to the GJT slides and correct the conditional sentences that they had judged ungrammatical (Appendix E for the GJT answer sheet and the correction sheet).

**Procedure**

Data were collected over three experimental sessions (see Figure 3). During the first meeting, learners signed the consent form, filled out the background questionnaire (biodata), and then carried out the WM task (OSPA). As the first meeting was conducted with groups of 10 to 15 learners, the OSPAN task was administered in groups of as many learners in a classroom using a computer and a computer projector equipped in a classroom at each institution/school, from which the target populations were recruited. A brief instruction session was provided before the span task. It took approximately 25 minutes for learners to finish the OSPAN task. On Day 8, each individual learner met the researcher and participated in the pretests, treatment, and
immediate posttests in the order shown in Figure 3. As mentioned earlier, the oral pretest and posttest were incorporated into the interactional activity that included the treatment. That is, the treatment was preceded by the oral pretest and followed by the oral posttest within the interactional task. Both tests were designed in the same format as in the treatment just with different drawings describing different incidents/events. Again, a brief instruction session was conducted on the GJT and the interactional task. Each GJT session on both targets took approximately 30 minutes. The interactional task involving the oral pretest, treatment, posttest took between 35 and 40 minutes. The posttests were followed by the task-set switching task, which the learner carried out for another 20 minutes. The learner was able to complete the entire tasks at the second meeting within 2 hours. Two weeks later, the delayed posttests (oral and GJT) were administered to the learner in a dyadic meeting with the researcher. The learner was also asked to fill out the exit questionnaire at the end of the session. The session took approximately one hour. As the present study was designed to investigate the role of WM and attention control in the effectiveness of corrective feedback, and no feedback was provided for the control group, neither the OSPAN task nor the task-set switching task was conducted by the control group.
Figure 3. Experimental Procedure.

| Day 1          | Consent form, Biodata, WM task (OSPAN) |
|               |                                        |
| Day 8         | GJT Pretest on both targets            |
|               | Interactional task (*Rashomon*)        |
|               | Oral Pretest                          |
|               | Treatment                              |
|               | Oral Posttest                          |
|               | GJT Posttest on both targets           |
|               | Task-set switching                     |
| Day 22        | Delayed posttests                     |
|               | (Oral followed by GJT on both targets) |
|               | Exit questionnaire                     |

Scoring

**OSPAN and task-set switching.** For the background or processing part of the OSPAN measure (i.e., the mathematical operation part of the OSPAN task), one point was given to each correct answer, and no point to each incorrect answer. Percentage scores were computed based on their performance on the processing part of the task. Regarding the recall part of the task, following Conway et al.’s (2005) recommendation on scoring WM span performance, the partial-credit unit scoring procedure was adopted (see also Friedman & Miyaki, 2005 for discussion on scoring methods). Conway et al. found, having reanalyzed Kane et al.’s (2004)
data, the highest internal consistency reliability via the partial-credit unit scoring procedure (see their review for more discussion on other alternatives such as partial-credit load scoring, all-or-nothing unit scoring, and all-or-nothing load scoring, all of which have been used in previous research). That is, one point was awarded to each correctly recalled set of items in its correct serial position within a set, and no point to an incorrectly recalled or not-recalled item. The number of correctly recalled items within each set was converted into a proportion-correct score; for instance, .6 was given when 3 items are correctly recalled in their correct serial positions in a set of 5. Then, an average proportion-correct score was calculated for each learner. As for the task-set switching task, the learner’s reaction time (RT) was measured both on the switch trials and nonswitch trials. The switch costs for each learner were calculated by subtracting the mean RT and error rate on the nonswitch trials from the corresponding values on the switch trials.

**Oral production/grammaticality judgment.** For the oral production task on the that-trace filter, one point was awarded to each correctly produced item and 0 to each incorrectly produced item in relation to the target. Regarding the second target, the past unreal conditional, each of the two clauses forming a past unreal conditional sentence (the if clause and the main/result clause) was scored as .5 if it was correctly produced according to the past unreal conditional rule. Therefore, one point was given to a conditional sentence if both the if clause and the main/result clause were grammatically correct, half a point to a sentence if only one of the clauses confirmed to the rule. No point was awarded to a sentence that failed to conform to the converting rule in either clause. Because learners were provided with 12 opportunities to produce sentences involving each target, the maximum possible score was 12.

With regard to the GJT on the that-trace filter, one point was awarded to each correct judgment for a total of 16 points. The selection of the choice *I don’t know* was interpreted as
having no knowledge about the *that*-trace filter, and hence no point was given.\(^3\) Also, in the case of ungrammatical sentences, one point was given only to a correct judgment accompanied by the provision of a grammatical version. As for the GJT on the past unreal conditional, one point was given to each correct judgment on each grammatical item. However, as in the oral production task on the same target, a clause-level scoring was conducted on learner performance on 8 ungrammatical items based on their provision of a grammatical version. Half a point was given to each correctly revised clause. Because there were 4 conditional sentences that were ungrammatical in both clauses, 2 sentences incorrect in the *if* clause only, and 2 sentences incorrect in the main/result clause only, the maximum possible score was 6 in the ungrammatical portion of the GJT; this made the total possible score in the GJT on the past unreal conditional 14 (8 points from grammatical items + 6 from ungrammatical items).

**Interrater Reliability**

The oral performance data were transcribed and scored according to the above-mentioned scoring criteria. Of the 83 learners in total, twenty-nine were randomly selected, and an independent rater scored their oral performance. Because the oral data were collected on two different occasions (the pretest and the posttest on the same day and the delayed posttest two weeks later), and some learners did not attend the delayed posttest session, two sets of oral performance data saved in the wav file format (one set for the pretest and the posttest and the other for the delayed posttest) were rendered to and scored by the independent rater (see Figure 4): 8 learners’ performance on the pretest and posttest only and 21 learners’ performance on the delayed posttest only (approximately 15% of the total oral performance data). The Cohen’s kappa was computed to assess the interrater reliability. The resulting kappa was .936, which
indicates the scoring of learner performance was fairly objective. Given this high interrater reliability, the remaining data were scored by the researcher alone.

Figure 4. Number of leaners selected from each group for interrater reliability.

![Chart showing number of learners selected from each group for interrater reliability.]

Analysis

It should be noted that data from those who showed their knowledge on the target(s) were excluded from relevant statistical analyses. To elaborate, if learners were able to provide correct versions for more than a half of the ungrammatical GJT items on the target(s), their data were excluded from analyses on the GJT performance. The same criterion was employed to determine whether to include/exclude data from related analyses of learner performance on the oral production test. Also, following the research tradition (e.g., Conway et al., 2005; Kane et al., 2004; Unsworth & Engle, 2007), only data from those whose math performance was higher than 85% in accuracy were included in relevant analyses involving WMC. Regarding the task switching data, incorrect trials, trials that immediately followed incorrect trials, and the first trial
of each block were excluded from the calculation of switch costs. Trials with RT of less than 100 ms or larger than 3000 ms were also discarded.

Just to clarify two statistical terms used in the results section, a mixed ANOVA and a repeated-measures ANOVA, a mixed ANOVA (also known as a repeated-measures ANOVA) is used to denote an ANOVA involving both a within-subjects variable (e.g., Time: pretest vs. posttest vs. delayed posttest) and a between-subjects variable (e.g., Group: recast vs. metalinguistic vs. control). A repeated-measures ANOVA also means an ANOVA with a within-subjects variable, but does not include a between-subjects variable. This is how they are differentiated in this dissertation.

Prior to every omnibus (principal/main) statistical test, learner performance on each pretest was always checked using a one-way ANOVA to see whether there was any significant difference in their knowledge/performance among the groups before the experiment. In one case, a Kruskal-Wallis test was adopted for this purpose. To test a main effect of Time and of Group in terms of learner performance, either a mixed ANOVA (when there was no significant between-groups difference in their performance on the pretest) or a mixed ANCOVA (when a significant between-groups difference was found in an ANOVA on the pretest) was performed as an omnibus test. For a significant between-subjects effect, if found in an omnibus test, a one-way ANOVA or ANCOVA was chosen as a follow-up statistical test to observe any possible significant difference among the groups at different points in time (e.g., posttest and/or delayed posttest). For a post hoc analysis, the Bonferroni correction was used to maintain the overall alpha level at .05 when comparing the groups to detect any statistical significance of mean differences. Each post hoc pairwise comparison was tested at a Bonferroni-corrected $p$ value of .017. When the assumption of equal variances in an ANOVA was violated, the Brown-
Forsythe procedure was selected to obtain an $F$ value and its significance. The Brown-Forsythe procedure was followed by Games-Howell post hoc comparisons. The Games-Howell test is recognized as a robust post hoc measure even when the assumptions of ANOVA are violated.

For a significant main effect of Time observed in a mixed ANOVA, a repeated-measures ANOVA (with a within-subjects variable only, that is, Time) was performed on the performance of each group on the pretest, posttest, and delayed posttest in an effort to see whether and how each group contributed to the significant effect of Time (and any effect of a treatment condition provided for each group, for that matter). A statistically significant effect of Time (pretest vs. posttest vs. delayed posttest) was accompanied by Bonferroni post hoc pairwise comparisons.

For mixed ANCOVAs, the Delaney-Maxwell method (Delaney & Maxwell, 1981) was utilized whereby each score in a covariate was adjusted based on the difference between the mean score and each individual score to prevent the covariate from affecting/altering any within-subjects effect of Time. The significance level was set at .05 for all statistical tests. All numeric values (e.g., $r$, $R^2$, $F$- and $p$-values) were rounded to three decimal places (e.g., $p = .652$).
Chapter IV: Results

Results are provided in this chapter in the following order: learner performance on the GJT and oral production involving the that-trace filter, learner performance on the GJT and the oral production involving the past unreal conditional, each of which is followed by a brief summary of results.

First, Cronbach’s alpha was used to measure the internal consistency reliability of items in each dependent variable measure (see Tables 5 and 6 for reliability coefficients). The overall results indicate the items in each measure were reliably consistent in measuring the same construct. Pearson bivariate correlations were computed between the two cognitive variables, WM and attention control. Somewhat surprisingly, WM measured via the OSPAN task was not correlated with either switch cost measure: \( r = .002 \) (\( p > .05 \)) between WM and the RT cost and \( r = -.182 \) (\( p > .05 \)) between WM and the error cost. This interesting phenomenon is discussed in the next chapter. However, as expected, a significant correlation with a medium effect size was found between the two switch cost measures, the RT cost and the error cost, \( r = .324 \) (\( p < .05 \)).

Table 5

<table>
<thead>
<tr>
<th>Type</th>
<th>That-Trace Filter</th>
<th>Past Unreal Conditional</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>.693</td>
<td>.866</td>
</tr>
<tr>
<td>B</td>
<td>.841</td>
<td>.842</td>
</tr>
<tr>
<td>C</td>
<td>.730</td>
<td>.841</td>
</tr>
</tbody>
</table>

Note. Cronbach’s alpha was used as a reliability measure based on their pretest performance.

Table 6

<table>
<thead>
<tr>
<th>Type</th>
<th>That-Trace Filter</th>
<th>Past Unreal Conditional</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>.971</td>
<td>.955</td>
</tr>
<tr>
<td>B</td>
<td>.949</td>
<td>.945</td>
</tr>
<tr>
<td>C</td>
<td>.855</td>
<td>.959</td>
</tr>
</tbody>
</table>

Note. Cronbach’s alpha was used as a reliability measure based on their pretest performance.
In order to make sure that the two experimental groups did not differ from each other in terms of cognitive capacities, three one-way ANOVAs were performed on the data collected from the task-set switching task (RT and error rate) and the OSPAN task. No statistically significant difference was found between the two experimental groups in learner performance on the OSPAN, \( F(1, 44) = 3.682, p > .05 \) (\( p = .062 \), also \( M = .828, SD = .084 \), and \( M = .875, SD = .084 \), for Groups 1 and 2, respectively) and on the task-set switching task, \( F(1, 43) = .725, p > .05 \) (\( p = .399 \)) on the RT cost, and \( F(1, 43) = .772, p > .05 \) (\( p = .385 \)) on the error cost (see Table 7 for means and standard deviations in the task-switching test). These nonsignificant findings indicate the two groups were more or less homogeneous with respect to their cognitive capacities reflected in the OSPAN and the task-set switching paradigm.

Table 7

**Descriptive Statistics on Results of Task-Set Switching**

<table>
<thead>
<tr>
<th>Switch</th>
<th>Nonswitch</th>
<th>Switch Costs</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>RT</strong></td>
<td><strong>Error (%)</strong></td>
<td><strong>RT</strong></td>
</tr>
<tr>
<td>Group 1 (( n = 22 ))</td>
<td>1347.57</td>
<td>8.50</td>
</tr>
<tr>
<td>Group 2 (( n = 23 ))</td>
<td>1286.50</td>
<td>7.20</td>
</tr>
</tbody>
</table>

*Note.* RT = mean reaction time in milliseconds. Error = mean error rate in percentage.

**That-Trace Filter**

**Grammaticality Judgment**

First of all, the data collected from those who participated in all three test sessions (pretest, posttest, and delayed posttest) were analyzed (see Table 8 for descriptive statistics and Figure 5 for a graphic presentation). A one-way ANOVA was conducted on the pretest performance to examine whether there was any significant between-groups difference before the treatment. The ANOVA found a significant between-groups difference, \( F(2, 52) = 3.714, p = .031, \eta_{p}^{2} = .125 \). Bonferroni post hoc pairwise comparisons showed that the metalinguistic group...
performed significantly better than the recast group on the pretest ($p < .017$), and that the other two pairwise comparisons (i.e., one between the recast and control groups and the other between the metalinguistic and control groups) were not statistically significant. However, due to this significant between-groups difference, a mixed ANCOVA was conducted on the GJT data with the pretest as a covariate, Time (posttest and delayed posttest) as a within-subjects variable, and Group (recast vs. metalinguistic vs. control) as a between-subjects variable.

Table 8

*Descriptive Statistics for Learner Performance on GJT on That-Trace Filter*

<table>
<thead>
<tr>
<th></th>
<th>Pretest</th>
<th>Posttest</th>
<th>Delayed</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$M$</td>
<td>$SD$</td>
<td>$M$</td>
</tr>
<tr>
<td>Recast ($n = 21$)</td>
<td>4.43</td>
<td>2.77</td>
<td>8.52</td>
</tr>
<tr>
<td>Meta ($n = 17$)</td>
<td>6.47</td>
<td>1.59</td>
<td>11.18</td>
</tr>
<tr>
<td>Control ($n = 17$)</td>
<td>5.47</td>
<td>2.27</td>
<td>6.24</td>
</tr>
</tbody>
</table>

*Figure 5. Results of GJT on that-trace filter.*

The ANCOVA showed no significant main effect of Time, $F (1, 51) = .687, p = .411, \eta^2_p = .013$, but a significant main effect of Group, $F (2, 51) = 5.637, p = .006, \eta^2_p = .181$. Neither the
Time x Group interaction nor the Time x Pretest (covariate) interaction was statistically significant: $F(2, 51) = 1.365, p = .265, \eta^2_p = .051$, $F(1, 51) = .171, p = .676, \eta^2_p = .003$, respectively. Further analyses were carried out on their performance due to the significant between-subjects (Group) difference. Two separate ANCOVAs were conducted, one on their posttest with the pretest as a covariate and the other on the delayed posttest with the pretest as a covariate. The ANCOVA on the posttest showed a significant difference among the groups, $F(2, 51) = 6.767, p < .05 (p = .002, \eta^2_p = .210)$. Bonferroni post hoc pairwise comparisons based on adjusted means after the pretest was statistically accounted for indicated that the metalinguistic group significantly outperformed the control group ($p < .017$), and but no significant difference was found either between the metalinguistic and recast groups or between the recast and control groups ($p > .017$). The ANCOVA on the delayed posttest with the pretest as a covariate also showed a significant between-subjects difference, $F(2, 51) = 3.806, p = .029, \eta^2_p = .130$. As in the posttest results, Bonferroni post hoc pairwise comparisons based on adjusted means found that the metalinguistic group performed significantly better than the control group ($p < .017$), but that there was no significant difference either between the metalinguistic and recast groups or between the recast and control groups ($p > .017$).

Both ANCOVAs provided evidence that the metalinguistic group performed significantly better than the control group although there was no difference between the two experimental groups. Nevertheless, these ANCOVAs provided no information on the within-subject variable that included the pretest performance of each group. Therefore, a repeated-measures ANOVA (with a within-subjects variable only, that it, the Time variable only) was conducted on the performance of each group on all three tests (pretest, posttest, and delayed posttest). As for the recast group, the repeated-measures ANOVA showed a significant effect of Time, $F(1.053,
21.066) = 10.634, \( p < .05 \) (\( p = .003, \eta^2_p = .347 \)); the Greenhouse-Geisser correction was made to
the degrees of freedom due to the violation of sphericity. Bonferroni pairwise comparisons
indicated that the recast group performed significantly better on both the posttest and the delayed
posttest than on the pretest, but that there was no significant difference between the two posttests.
A similar trend was observed in the performance of the metalinguistic group. There was a
significant effect of Time, \( F(2, 32) = 15.253, p < .05 \) (\( p = .000, \eta^2_p = .488 \)). The metalinguistic
group performed significantly better on the posttest and the delayed posttest than on the pretest
(\( p < .017 \)), but there was no significant difference between the posttest and the delayed posttest
(\( p > .017 \)). With regard to the control group, no significant effect of Time was evidenced despite
a significant trend, \( F(2, 32) = 3.195, p > .05 \) (\( p = .054, \eta^2_p = .166 \)).

Another set of statistical analyses was performed on the data collected only from the
pretest and the posttest because some learners did not show up for the delayed session. This was
designed not to miss any potentially important information (see Table 9 and Figure 6).

Table 9

<table>
<thead>
<tr>
<th></th>
<th>Pretest</th>
<th></th>
<th>Posttest</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( M )</td>
<td>( SD )</td>
<td>( M )</td>
<td>( SD )</td>
</tr>
<tr>
<td>Recast (( n = 23 ))</td>
<td>4.65</td>
<td>2.74</td>
<td>8.43</td>
<td>4.33</td>
</tr>
<tr>
<td>Meta (( n = 20 ))</td>
<td>6.20</td>
<td>1.67</td>
<td>11.05</td>
<td>3.99</td>
</tr>
<tr>
<td>Control (( n = 32 ))</td>
<td>5.47</td>
<td>2.23</td>
<td>5.69</td>
<td>2.98</td>
</tr>
</tbody>
</table>

* \( p < .05 \).
Figure 6. Results of GJT on that-trace filter (pretest and posttest only).

A one-way ANOVA was conducted on the pretest to see whether the groups differed in their performance on the pretest. Unlike the ANOVA on the pretest when all three sessions were included, no significant difference among the groups was observed in the pretest, $F(2, 72) = 2.494, p = .090, \eta^2_p = .065$. A mixed ANOVA was thus performed with Time (pre and post only) as a within-subjects variable and Group (recast vs. metalinguistic vs. control) as a between-subjects variable. The mixed ANOVA found a significant main effect of Time, $F(1, 72) = 36.357, p = .000, \eta^2_p = .336$, of Group, $F(2, 72) = 11.185, p = .000, \eta^2_p = .237$, and of the interaction between Time and Group, $F(2, 72), p = .000, \eta^2_p = .202$.

To further analyze the significant main effect of Time (pretest to posttest), three paired-sample $t$-tests were conducted on the pretest and posttest results of the groups. A significant improvement from the pretest to the posttest was evidenced in the performance of both experimental groups, $t(22) = -3.343, p = .003, d = 1.04$ and $t(19) = -5.314, p = .000, d = 1.59$ for the recast and metalinguistic groups, respectively. Almost no pre-to-post improvement was
observed in the performance of the control group, \( t(31) = -.412, p = .683, d = .08 \). Regarding the significant main effect of Group, because the ANOVA on the pretest showed no statistical significance, a one-way ANOVA on the posttest was conducted. There was a significant difference among the groups, \( F(2, 72) = 13.149, p = .000, \eta_p^2 = .268 \). Bonferroni post hoc pairwise comparisons indicated that both experimental groups significantly outperformed the control group \((p < .017)\), and that no significant difference was observed between the two experimental groups. The significant interaction effect stemmed mainly from the fact that the control group performed somewhat better, albeit negligible, than the recast group in the pretest, but the recast group significantly outperformed the control group in the posttest (see Figure 6).

Lastly, as in Goo (in press), to detect any significant difference among the groups as to their pre-to-post development, normalized gain scores (Hake, 1998) were calculated in terms of proportion by dividing the number of posttest items correct minus the number of pretest items correct by the number of pretest items incorrect, \((M = .289 \text{ and } SD = .385 \text{ for the recast group, } M = .492 \text{ and } SD = .423 \text{ for the metalinguistic group, and } M = -.003 \text{ and } SD = .331 \text{ for the control group})\). Similarly, pre-to-delayed gain scores were calculated by dividing the number of items correct in the delayed posttest minus the number of items correct in the pretest by the number of items incorrect in the pretest \((M = .296 \text{ and } SD = .382 \text{ for the recast group, } M = .414 \text{ and } SD = .395 \text{ for the metalinguistic group, and } M = .080 \text{ and } SD = .199 \text{ for the control group})\). Statistical analyses based on these normalized gain scores, described below and throughout this chapter, were designed to examine the extent to which learner development differs due to their treatment conditions.

A one-way ANOVA performed on the pre-to-post gain score data showed a significant difference among the groups, \( F(2, 72) = 11.368, p = .000, \eta_p^2 = .240 \). Bonferroni post hoc
pairwise comparisons indicated that both experimental groups obtained significantly higher pre-to-post gain scores than the control group \( (p < .017) \), but that there was no significant difference between the two experimental groups. As for the pre-to-delayed development on the *that*-trace filter, a significant between-groups difference was found, \( F (2, 52) = 4.201, p = .020, \eta_p^2 = .139 \). The metalinguistic group significantly outperformed the control group \( (p < .017) \), but no significant difference was observed either between the two experimental groups or between the recast and control groups \( (p > .017) \), which reflects the statistical results obtained in the earlier analyses of the data collected from those who participated in all three tests.

Concerning the role of WM and/or attention control in L2 learning through corrective feedback, first and foremost, simple linear regression analyses were carried out to make a statistical decision on whether ANCOVAs with a cognitive variable as a covariate should be used in comparing just the two experimental groups; the cognitive tasks were administered only to the two experimental groups. Each of the two posttests (date from those who participated in all three test sessions and whose accuracy in the math task was above 85%) was regressed on each cognitive variable measure. Neither WM nor attention control significantly predicted the overall performance of the two experimental groups on the posttest and the delayed posttest (see Table F1 in Appendix F).

Furthermore, because different sample sizes were involved in the pretest-and-posttest-only analyses described earlier, three simple regression analyses were conducted to see whether an ANCOVA with a cognitive factor as a predictor variable was necessary to compare the two experimental groups. None of the cognitive factors was found to have played any significant predictive role in the overall performance of the experimental groups on the posttest (see Table F2 in Appendix F). Also, the pre-to-post gain scores of the experimental groups were not found
to be significantly associated with either cognitive factor. Similarly, no significant results were observed with regard to the pre-to-delayed development reflected in normalized gain scores (see Table F3 in Appendix F).

As a further step to investigate the potential impacts of the cognitive variables (WM and attention control) on the effectiveness of each feedback type in the acquisition of knowledge on the that-trace filter, two sets of simple linear regression analyses were conducted, one on pre-to-post gain scores and the other on pre-to-delayed gain scores (for each feedback condition). No statistically significant evidence was obtained that indicated any mediating role(s) of WM and attention control in the observed effects of recasts or metalinguistic feedback as reflected in gain scores (see Tables F4 and F5 in Appendix F).

**Summary of GJT Performance on English That-Trace Filter**

When the data collected from those who participated in all three test sessions were analyzed, metalinguistic feedback seemed to be slightly more effective than recasts, given only the metalinguistic group significantly outperformed the control group despite no significant difference between the two experimental groups. However, when only the pretest and the posttest (more data) were considered, the recast group as well as the metalinguistic group performed significantly better than the control group. These patterns of performance differences were also observed in the statistical results of pre-to-post and pre-to-delayed gain scores. That is, while it is only the metalinguistic group that outperformed the control group in pre-to-delayed gain scores, both the recast and metalinguistic groups performed significantly better than the control group. Given these results, it appears that recasts and metalinguistic feedback were equally beneficial for the acquisition of knowledge on the English *that*-trace filter. Also, neither
WM nor attention control was found to be associated with the observed beneficial effects of recasts or metalinguistic feedback.

**Oral Production**

Table 10 shows descriptive statistics for learner performance on the oral production task on the English *that*-trace filter (see also Figure 7). Because the assumptions of normality and homogeneity of variances were clearly violated in several cases, and different sample sizes were used for the groups, a Kruskal-Wallis test was conducted on the oral pretest. A significant between-groups difference in their pretest performance was obtained, $\chi^2 (2) = 7.782, p < .05$ ($p = .020$). The homogeneity of variance assumption was also violated in the performance of the groups on the posttest. Thus, an omnibus test (i.e., a mixed ANCOVA) was not performed on the data collected only from those who participated in all three sessions. This decision was due to a limited set of nonparametric statistical tests. Rather, a repeated-measures ANOVA only with a within-subjects variable “Time” (pretest, posttest, delayed posttest) on the performance of each group was carried out, and its results are reported in the following.

Table 10

<table>
<thead>
<tr>
<th></th>
<th>Pretest</th>
<th>Posttest</th>
<th>Delayed</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$M$</td>
<td>$SD$</td>
<td>$M$</td>
</tr>
<tr>
<td>Recast ($n = 19$)</td>
<td>.47</td>
<td>.91</td>
<td>7.58</td>
</tr>
<tr>
<td>Meta ($n = 16$)</td>
<td>.94</td>
<td>1.57</td>
<td>12.00</td>
</tr>
<tr>
<td>Control ($n = 12$)</td>
<td>.00</td>
<td>.00</td>
<td>.17</td>
</tr>
</tbody>
</table>
Figure 7. Learner performance on oral production task on that-trace filter.

The repeated-measures ANOVA on the performance of the recast group showed a significant main effect of Time, $F(2, 36) = 21.059, p < .05$ ($p = .000, \eta^2_p = .539$). Bonferroni pairwise comparisons showed that the recast group performed significantly better on the posttest and the delayed posttest than on the pretest ($p < .017$), but there was no significant difference between the two posttests. A significant main effect of Time was also observed in the performance of the metalinguistic group, $F(1.225, 18.381) = 94.045, p < .05$ ($p = .000, \eta^2_p = .862$); the Greenhouse-Geisser adjustment was used for the degrees of freedom due to the violation of sphericity. As in the case of the recast group, the metalinguistic group performed significantly better on the two posttests than on the pretest ($p < .017$), and no significant difference was found between the two posttests. Unsurprisingly, no significant effect of Time was evidenced in the performance of the control group, $F(1.069, 11.755) = .975, p > .05$ ($p = .350, \eta^2_p = .081$); the Greenhouse-Geisser adjustment was used for the degrees of freedom due to the violation of sphericity.
As mentioned earlier with respect to the GJT, some learners did not participate in the delayed session. Thus, learner performance only on the pretest and the posttest was analyzed. The results of paired-samples t-tests and corresponding effect sizes as well as descriptive statistics for their performance on the oral pretest and posttest sessions are shown in Table 11.

Table 11

| Results of Oral Performance on That-trace (Pretest and Posttest Only) |
|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|
|                         | Pretest                  | Posttest                 | T                     | df | P | d |
| Recast (n = 21)         | .43                      | .87                      | 7.76                   | 5.63 | -5.719* | 20 | .000 | 1.82 |
| Meta (n = 19)           | .79                      | 1.48                     | 11.37                  | 2.75 | -15.632* | 18 | .000 | 4.79 |
| Control (n = 23)        | .09                      | .42                      | .43                    | .99 | -1.785 | 22 | .088 | .45 |

*p < .05

Figure 8. Results of oral production task on that-trace filter (pretest and posttest only)

As shown in Table 11 and Figure 8, both experimental groups performed significantly better on the oral posttest than on the oral pretest. Next, learner performance on the oral pretest was analyzed using the Brown-Forsythe procedure due to the violation of the assumption of equal variances. It showed no significant difference among the groups, $F(2, 32.191) = 2.409, p > .05$ ($p = .106$). Their performance on the oral posttest was also submitted to the Brown-Forsythe
procedure due again to the violation of the homogeneity of variance assumption. A significant between-groups difference was found in their performance on the posttest, $F(2, 30.615) = 48.926, p < .05 (p = .000)$. The Games-Howell post hoc test was used to compare the performance of the groups. The results of three pairwise comparisons indicated that the metalinguistic group significantly outperformed the recast and control groups ($ps = .036$ and .000, respectively), and that the recast group performed significantly better than the control group ($p = .000$).

Another Brown-Forsythe procedure was conducted to detect any significant difference among the groups in their pre-to-post gain scores ($M = .624$ and $SD = .503$ for the recast group, $M = .947$ and $SD = .229$ for the metalinguistic group, and $M = .029$ and $SD = .078$ for the control group). As with the oral posttest, a significant between-groups difference was observed, $F(2, 29.248) = 44.149, p < .05 (p = .000)$. According to the Games-Howell post hoc tests, the metalinguistic group showed significantly higher pre-to-post gain scores than the recast group ($p = .033$) as well as the control group ($p = .000$), and the recast group obtained significantly higher pre-to-post gain scores than the control group ($p = .000$). As for pre-to-delayed gain scores ($M = .556$ and $SD = .505$ for the recast group, $M = .828$ and $SD = .346$ for the metalinguistic group, and $M = .069$ and $SD = .216$ for the control group), a one-way ANOVA was conducted on the data. The ANOVA showed a significant mean difference among the groups, $F(2, 44) = 12.674, p < .05 (p = .000, \eta^2_p = .366)$. Both experimental groups improved significantly more than did the control group ($p < .017$). No significant difference was found between the two experimental groups.

As in the analyses of the GJT performance, several regression analyses were conducted to detect any potential involvement of WM and attention control in learner performance and
treatment effects. Data collected only from the two experimental groups were regressed on WM and attention control. However, because the data from those who participated in all three tests were not analyzed in a general linear model due to the violations of the equal variance and normality assumptions and a significant between-groups difference in the pretest, regression analyses were not performed on those data. WM and attention control were found to be unrelated to learner performance on the oral task (see Tables in Appendix G).

**Summary of Oral Performance on English That-Trace Filter**

Treatment effects were clearly evidenced in the performance of the two experimental groups on the posttest and delayed posttest. With the data from the pretest and the posttest only, the metalinguistic group significantly outperformed the recast group as well as the control group on the posttest, and the recast group performed significantly better than the control group. A similar trend was observed in terms of pre-to-post gain scores, that is, the metalinguistic group obtained significantly higher gain scores than the other two groups. As for pre-to-delayed gain scores, both experimental groups showed significantly higher gain scores than the control group, and no significant difference was found between the two experimental groups. Overall, metalinguistic feedback was more effective than recasts in oral production involving the that-trace filter. Also, WM and attention control were not found to be associated with the learners’ oral performance (see Table 12 for a summary of overall results on the English that-trace filter).
Table 12

*Summary of Results on English That-Trace Filter*

<table>
<thead>
<tr>
<th></th>
<th><strong>GJT Performance</strong></th>
<th><strong>Oral Performance</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>All 3 Tests Involved</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Posttest</td>
<td>Meta &gt; Control,</td>
<td>Posttest, Delayed &gt; Pretest</td>
</tr>
<tr>
<td></td>
<td>Meta = Recast,</td>
<td>Posttest = Delayed</td>
</tr>
<tr>
<td></td>
<td>Recast = Control</td>
<td></td>
</tr>
<tr>
<td>Delayed posttest</td>
<td>Meta &gt; Control,</td>
<td>Posttest, Delayed &gt; Pretest</td>
</tr>
<tr>
<td></td>
<td>Meta = Recast,</td>
<td>Posttest = Delayed</td>
</tr>
<tr>
<td></td>
<td>Recast = Control</td>
<td></td>
</tr>
<tr>
<td>Recast</td>
<td>Posttest, Delayed &gt; Pretest</td>
<td>Posttest, Delayed &gt; Pretest</td>
</tr>
<tr>
<td></td>
<td>Posttest = Delayed</td>
<td>Posttest = Delayed</td>
</tr>
<tr>
<td>Metalinguistic</td>
<td>Posttest, Delayed &gt; Pretest</td>
<td>Posttest, Delayed &gt; Pretest</td>
</tr>
<tr>
<td></td>
<td>Posttest = Delayed</td>
<td>Posttest = Delayed</td>
</tr>
<tr>
<td>Control</td>
<td>Pretest = Posttest = Delayed</td>
<td>Pretest = Posttest = Delayed</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Two Tests Involved</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Posttest</td>
<td>Recast, Meta &gt; Control</td>
<td>Meta &gt; Recast &gt; Control</td>
</tr>
<tr>
<td></td>
<td>Recast = Meta</td>
<td></td>
</tr>
<tr>
<td>Recast</td>
<td>Posttest &gt; Pretest</td>
<td>Posttest &gt; Pretest</td>
</tr>
<tr>
<td>Metalinguistic</td>
<td>Posttest &gt; Pretest</td>
<td>Posttest &gt; Pretest</td>
</tr>
<tr>
<td>Control</td>
<td>Posttest = Pretest</td>
<td>Posttest = Pretest</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pre-to-Post Gain</td>
<td>Recast, Meta &gt; Control</td>
<td>Meta &gt; Recast &gt; Control</td>
</tr>
<tr>
<td></td>
<td>Recast = Meta</td>
<td></td>
</tr>
<tr>
<td>Pre-to-Delayed Gain</td>
<td>Meta &gt; Control</td>
<td>Recast, Meta &gt; Control</td>
</tr>
<tr>
<td></td>
<td>Meta = Recast,</td>
<td>Recast = Meta</td>
</tr>
<tr>
<td></td>
<td>Recast = Control</td>
<td></td>
</tr>
</tbody>
</table>

**Cognitive Factors**

<table>
<thead>
<tr>
<th></th>
<th><strong>GJT Performance</strong></th>
<th><strong>Oral Performance</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>WMC</td>
<td>No significant relationship</td>
<td>No significant relationship</td>
</tr>
<tr>
<td>RT Cost</td>
<td>No significant relationship</td>
<td>No significant relationship</td>
</tr>
<tr>
<td>Error Cost</td>
<td>No significant relationship</td>
<td>No significant relationship</td>
</tr>
</tbody>
</table>

*Note.* “=” means no statistically significant difference.
Past Unreal Conditional

Grammaticality Judgment

Turning to the second linguistic target, learner performance on the GJT's on the past unreal conditional, using data collected from those who participated in all three test sessions (pretest, posttest, and delayed posttest), was first analyzed (see Table 13 for descriptive statistics and Figure 9 for a graphic presentation). A one-way ANOVA was conducted on their pretest performance to examine whether there was any significant between-groups difference before the actual experiment. No significant between-groups difference was found, $F(2, 40) = .126, p = .882, \eta^2_p = .006$.

Table 13

<table>
<thead>
<tr>
<th></th>
<th>Pretest</th>
<th>Posttest</th>
<th>Delayed</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$M$</td>
<td>$SD$</td>
<td>$M$</td>
</tr>
<tr>
<td>Recast ($n = 17$)</td>
<td>7.24</td>
<td>2.29</td>
<td>10.03</td>
</tr>
<tr>
<td>Meta ($n = 15$)</td>
<td>7.30</td>
<td>2.46</td>
<td>9.73</td>
</tr>
<tr>
<td>Control ($n = 11$)</td>
<td>6.86</td>
<td>2.15</td>
<td>8.45</td>
</tr>
</tbody>
</table>

Figure 9. Results of GJT on past unreal conditional.
Following this, a mixed ANOVA was performed with Time as a within-subjects variable (pretest vs. posttest vs. delayed posttest) and Group as a between-subjects variable (recast vs. metalinguistic feedback vs. control). The only significant finding was observed on the within-subjects variable, Time, \( F(1.568, 62.722) = 20.225, p < .05 \) \((p = .000)\); Greenhouse-Geisser correction was made to the degrees of freedom due to the violation of sphericity.

As a further step to examine the locus of the significant within-subjects effect, a repeated-measures ANOVA was carried out on the performance of each group on the pretest, posttest, and delayed posttest. First, a significant difference was found in the performance of the recasts group, \( F(1.286, 20.576) = 12.078, p = .001, \eta_p^2 = .430; \) Greenhouse-Geisser correction was made to the degrees of freedom. Three Bonferroni post hoc pairwise comparisons indicated that the recast group performed significantly better on the posttest and delayed posttest than on the pretest \((p < .017)\), and no significant difference was found between the posttest and the delayed posttest. Also, a significant difference in the performance of the metalinguistic group was observed, \( F(1.421, 19.898) = 5.898, p = .016, \eta_p^2 = .296; \) again Greenhouse-Geisser correction was made to the degrees of freedom. According to the results of three Bonferroni post hoc pairwise comparisons, the metalinguistic group performed significantly better on the posttest than on the pretest \((p < .017)\). However, no significant difference was found either between the pretest and the delayed posttest or between the posttest and the delayed posttest \((ps > .017)\). As for the control group, although the within-subjects variable (Time) was significant \( F(2, 20) = 5.640, p = .011, \eta_p^2 = .361 \) \((p < .05)\), no significant difference was evidenced in any of three Bonferroni post hoc pairwise comparisons \((p > .017)\).

Therefore, the significant main effect of Time evidenced in the mixed ANOVA earlier seems to have stemmed, for the most part, from the performance of the experimental groups.
Again, the data only on the pretest and the posttest were analyzed, and their results are described in the following (see Table 14 and Figure 10).

Table 14

<table>
<thead>
<tr>
<th></th>
<th>Pretest</th>
<th></th>
<th>Posttest</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>Recast (n = 22)</td>
<td>6.20</td>
<td>2.87</td>
<td>9.59</td>
<td>2.55</td>
</tr>
<tr>
<td>Meta (n = 17)</td>
<td>7.26</td>
<td>2.31</td>
<td>9.56</td>
<td>2.03</td>
</tr>
<tr>
<td>Control (n = 22)</td>
<td>6.61</td>
<td>1.84</td>
<td>7.73</td>
<td>2.00</td>
</tr>
</tbody>
</table>

Figure 10. Results of GJT on past unreal conditional (pretest and posttest only).

The ANOVA on the pretest did not find any significant between-groups difference, $F(2, 58) = .953, p = .392, \eta^2 = .032$. Thus, a 2 x 3 mixed ANOVA was conducted on the learners’ pretest and posttest performance with Time (pretest vs. posttest) as a within-subjects variable and Group (recast vs. metalinguistic vs. control) as a between-subjects variable. The ANOVA found a significant main effect of Time, $F(1, 58) = 55.883, p = .000, \eta^2 = .491$, and of Time x Group, $F(2, 58) = 5.152, p = .009, \eta^2 = .151$, but a nonsignificant effect of Group, $F(2, 58) = 1.941, p = .153, \eta^2 = .063$. 

125
To further investigate the significant effect of Time, three paired-samples $t$-tests were conducted on the results of the pretest and the posttest (one paired-samples $t$-test for each group). All three groups performed significantly better on the posttest than on the pretest in terms of accuracy in the GJT with both experimental groups showing large effect sizes and the control group showing a medium effect size: $t(21) = -5.109, p = .000, d = 1.25$ for the recast group, $t(16) = -5.245, p = .000, d = 1.06$ for the metalinguistic group, and $t(21) = -2.970, p = .007, d = .58$ for the control group. The significant Time x Group interaction seems attributable to the improved performance of the recast group. That is, the other two groups outperformed the recast group in the pretest, but the recast group performed better, albeit not statistically significant, than the other two on the posttest as shown in Figure 10.

Next, pre-to-post and pre-to-delayed gain scores were analyzed. A one-way ANOVA was first performed on the pre-to-post development data ($M = .403$ and $SD = .330$ for the recast group, $M = .353$ and $SD = .272$ for the metalinguistic group, and $M = .142$ and $SD = .251$ for the control group). There was a significant difference among the groups in their pre-to-post gain scores, $F(2, 58) = 5.020, p = .010, \eta_p^2 = .148$. Three Bonferroni post hoc pairwise comparisons indicated that the recast group surpassed the control group in the pre-to-post development at a statistically significant level ($p < .017$), but the pre-to-post development of the metalinguistic group was not significantly larger than that of the control group ($p > .017$). Also, no significant difference was found between the two experimental groups either ($p > .017$). However, an ANOVA on their pre-to-delayed gain scores ($M = .345$ and $SD = .444$ for the recast group, $M = .241$ and $SD = .546$ for the metalinguistic group, and $M = .287$ and $SD = .280$ for the control group) showed no significant between-groups difference, $F(2, 40) = .213, p = .809, \eta_p^2 = .011$. 

126
With respect to the role of WM and attention control in the efficacy of corrective feedback, learner performance on the GJT was regressed on the results of each cognitive variable measure. WM and attention control were not found to be associated with the observed effects of corrective feedback (see Tables in Appendix H).

**Summary of GJT Performance on Past Unreal Conditional**

In sum, when the data from all three sessions were considered, recasts had a longer-lasting effect on the learning of the English past unreal conditional. That is, the recast group performed significantly better on both posttests than on the pretest, whereas the metalinguistic group performed more or less similarly on the pretest and the delayed posttest, showing a sign of receding impacts of metalinguistic feedback observed in the performance on the posttest. The control group, however, did not perform significantly better on the two posttests than on the pretest; no significant difference was identified in any pairwise comparisons at a Bonferroni-corrected $p$ value of .017. An additional ANOVA on the data only from the pretest and the posttest showed that the groups did not differ from one another in their performance on the posttest. Nonetheless, the recast group obtained significantly higher gain scores than the control group. No significant difference was found in pre-to-post gain scores between the metalinguistic group and the control group or between the two experimental groups. The groups did not significantly differ from one another in terms of pre-to-delayed gain scores. Overall, it appears that recasts were slightly more effective than metalinguistic feedback at facilitating the acquisition of knowledge on the English past unreal conditional. No statistically significant evidence was obtained which indicated a significant role of WM or attention control in the acquisition of knowledge on the past unreal conditional through recasts or metalinguistic feedback.
Oral Production

Displayed in Table 15 and Figure 11 are descriptive statistics for learner performance on the oral production task. A one-way ANOVA was performed on the pretest in order to check whether the groups had begun the experiment on an equal footing.

Table 15

<table>
<thead>
<tr>
<th></th>
<th>Pretest</th>
<th>Posttest</th>
<th>Delayed</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$M$</td>
<td>$SD$</td>
<td>$M$</td>
</tr>
<tr>
<td>Recast ($n = 16$)</td>
<td>1.44</td>
<td>2.06</td>
<td>10.22</td>
</tr>
<tr>
<td>Meta ($n = 16$)</td>
<td>2.06</td>
<td>2.41</td>
<td>9.09</td>
</tr>
<tr>
<td>Control ($n = 11$)</td>
<td>.23</td>
<td>.34</td>
<td>.45</td>
</tr>
</tbody>
</table>

Figure 11. Learner performance on oral production task on past unreal conditional.

Due to the violation of the assumption of equal variances shown in Levene’s statistic ($p < .05$), learner performance on the pretest was analyzed using the Brown-Forsythe procedure. A significant between-groups difference was observed in the procedure, $F (2, 30.063) = 3.452, p < .05 (p = .045)$. Games-Howell post hoc comparisons showed the metalinguistic group
significantly outperformed the control group ($p < .05$), but no significant difference was found either between the recast and metalinguistic groups or between the recast and control groups ($p > .05$). Because of the significant between-groups difference in their performance on the pretest, a mixed ANCOVA was conducted with Time as a within-subjects variable (posttest vs. delayed posttest), Group as a between-subjects variable (recast vs. metalinguistic vs. control), and Pretest as a covariate.

The mixed ANCOVA showed a significant main effect of Group, $F (2, 39) = 27.654, p = .000, \eta^2_p = .586$, and of Time x Group interaction, $F (2, 39) = 5.912, p = .006, \eta^2_p = .233$. There was no significant main effect of Time, $F (1, 39) = 2.273, p = .140, \eta^2_p = .055$. That is, the learners’ performance on the posttest did not significantly differ from their performance on the delayed posttest after their pretest performance was statistically controlled for. With regard to the significant Group effect, Bonferroni post hoc pairwise comparisons were conducted based on adjusted scores/means. Both experimental groups performed significantly better than the control group ($p < .017$), but no significant difference was found between the two groups ($p > .017$). To obtain further information on the significant effect of Group, two separate ANCOVAs were performed: one on the posttest with the pretest as a covariate and the other on the delayed posttest with the pretest as a covariate.

The ANCOVA on the posttest with the pretest as a covariate showed a significant between-groups effect, $F (2, 39) = 75.841, p < .05, p = .000, \eta^2_p = .795$. Three Bonferroni post hoc pairwise comparisons based on adjusted means indicated that both experimental groups performed significantly better than the control group ($p < .017$), and there was no significant difference between the two experimental groups ($p > .017$). Observed in the ANCOVA on the delayed posttest with the pretest as a covariate was a significant between-groups difference, $F (2,
39) = 5.603, \( p < .05 \) \( p = .007 \), \( \eta_p^2 = .223 \). Three Bonferroni pairwise comparisons based on adjusted means, similarly to the ANCOVA on the posttest, indicated that both experimental groups significantly outperformed the control group (\( p < .017 \)), but that there was no significant difference between the two experimental conditions (\( p > .017 \)).

The significant Time x Group interaction obtained in the mixed ANCOVA earlier may have been attributed to the performance of the two experimental groups. The recast group outperformed the metalinguistic group on the posttest, whereas the metalinguistic group performed better, although negligible, than the recast group on the delayed posttest (see Table 15 and Figure 11 for a similar change from the posttest and the delayed posttest based on unadjusted means and standard deviations).

The mixed ANCOVA described above does not provide information on the performance of each group at all three different points in time, and the ANCOVAs were performed because of the significant between-groups difference on the pretest. Therefore, a repeated-measures ANOVA (only with a within-subjects variable, Time) was conducted on the performance of each group (pretest, posttest, and delayed posttest). As for the recast group, a significant within-subjects (Time) difference was obtained, \( F (1.337, 20.060) = 50.727, \ p < .05, \ p = .000, \ \eta_p^2 = .772 \) (Greenhouse-Geisser correction was made to the degrees of freedom due to the violation of sphericity). The recast group performed significantly better on the posttest than on the pretest or the delayed posttest (\( p < .017 \)). Learner performance on the delayed posttest was significantly better than on the pretest (\( p < .017 \)). A significant within-subjects (Time) effect was also found in the performance of the metalinguistic group, \( F (2, 30) = 43.792, \ p < .05, \ p = .000, \ \eta_p^2 = .745 \). The oral production of the metalinguistic group was significantly more accurate in the two posttests than in the pretest (\( p < .017 \)), but no significant difference in accuracy was found
between the two posttests ($p > .017$). Unlike the two treatment groups, the control group performed similarly on all three tests, $F(1.097, 10.971) = 2.504, p > .05, p = .141, \eta_p^2 = .200$ (Greenhouse-Geisser correction to the degrees of freedom).

The data only from the pretest and the posttest were also analyzed (see Table 16 and Figure 12). A one-way ANOVA conducted on the pretest showed no significant between-groups difference, $F(2, 58) = 1.632, p = .204, \eta_p^2 = .053$. Thus, a 2 x 3 mixed ANOVA was carried out with Time as a within-subjects variable (pretest and posttest) and Group as a between-subjects variable (recast vs. metalinguistic vs. control).

Table 16

<table>
<thead>
<tr>
<th>Results of Oral Performance on Conditional (Pretest and Posttest Only)</th>
<th>Pretest</th>
<th>Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$M$</td>
<td>$SD$</td>
</tr>
<tr>
<td>Recast ($n = 21$)</td>
<td>1.67</td>
<td>2.05</td>
</tr>
<tr>
<td>Meta ($n = 18$)</td>
<td>1.94</td>
<td>2.30</td>
</tr>
<tr>
<td>Control ($n = 22$)</td>
<td>.91</td>
<td>1.30</td>
</tr>
</tbody>
</table>

*p < .05

Figure 12. Oral performance on conditional (pretest and posttest only)
The mixed ANOVA showed a significant main effect of Time, $F(1, 58) = 342.557, p = .000, \eta^2_p = .855$, of Group, $F(2, 58) = 41.985, p = .000, \eta^2_p = .591$, and of Time x Group interaction, $F(2, 58) = 76.540, p = .000, \eta^2_p = .725$. As regards the significant Time effect (posttest > pretest), paired-samples t-tests were conducted to examine the contribution that each group made to the significant effect of Time obtained in the mixed ANOVA. The recast group produced significantly more accurate utterances in the posttest than in the pretest, showing a large effect size, $t(20) = -15.267, p = .000, d = 3.68$. A similar trend was evidenced in the performance of the metalinguistic group with a large effect size, $t(17) = -10.712, p = .000, d = 2.80$. No significant difference between the pretest and the posttest was found for the control group, $t(21) = -1.870, p = .076, d = .26$.

With regard to the significant effect of Group, learner performance on the posttest was submitted to a one-way ANOVA. A significant between-groups difference was obtained, $F(2, 58) = 83.967, p = .000, \eta^2_p = .743$. Three Bonferroni post hoc pairwise comparisons showed that the two experimental groups produced significantly more accurate utterances than did the control group ($p < .017$), but no significant difference between the two experimental groups was witnessed ($p > .017$). Given the posttest contributed to the significant effect of Group, and considering there was no significant difference among the groups in the pretest performance, the significant Group effect was attributable mainly to the improved performance of the two experimental groups.

The significant Time x Group interaction seems to be a result of the performance of the two experimental groups. The metalinguistic group performed somewhat better than the recast group in the pretest, however, the latter performed better than the former after the experimental treatment. Also, a significant increase in accuracy in the performance of the two experimental
conditions likely contributed to the significant interaction effect in that the significant between-groups difference was significantly larger in the posttest performance than in the pretest performance, obviously, between the control group and the experimental groups.

The pre-to-post improved performance computed in terms of normalized gain scores ($M = .783$ and $SD = .203$ for the recast group, $M = .721$ and $SD = .236$ for the metalinguistic group, and $M = .044$ and $SD = .107$ for the control group) was also submitted to a one-way ANOVA. A significant between-groups difference was observed in the ANOVA, $F (2, 58) = 102.688, p = .000, \eta_p^2 = .780$. Three Bonferroni post hoc comparisons showed a significantly larger increase in accuracy in the oral performance of both experimental groups, compared to the performance of the control group ($p < .017$). No significant difference was observed between the two experimental groups in terms of their pre-to-post development ($p > .017$).

Also analyzed was the pre-to-delayed development in terms of accuracy in the oral production task ($M = .579$ and $SD = .370$ for the recast group, $M = .640$ and $SD = .384$ for the metalinguistic group, and $M = .130$ and $SD = .243$ for the control group). Because there was a significant between-groups difference in learner performance on the pretest when the delayed posttest was included in earlier analyses, an ANCOVA with the pretest as a covariate was conducted on pre-to-delayed gain scores. A significant between-groups difference was found in the ANCOVA, $F (2, 39) = 4.965, p = .012, \eta_p^2 = .203$. Bonferroni pairwise comparisons based on adjusted means after the pretest performance was statistically controlled for showed that both experimental groups significantly outperformed the control group ($p < .017$), but the experimental groups did not differ from each other ($p > .017$).

As in previous analyses, several regression analyses were conducted in order to detect any potentially mediating role of WM and attention control in the efficacy of recasts and/or
metalinguistic feedback. Thus, only the performance of the experimental groups was included in these regression analyses. Learner performance on the posttest and the delayed posttest was regressed on WM and attention control (both RT and error costs). The data from those who participated in all test sessions were first analyzed. Three simple linear regression analyses were carried out to examine the impacts of WM and attention control on the overall performance of the experimental groups on each posttest.

WM significantly predicted the overall performance of the experimental groups on the oral posttest, showing a medium effect size, $F(1, 28) = 5.327, p = .029, R = .400, R^2 = .160$, adjusted $R^2 = .130$ (see also Table I1 in Appendix I). The adjusted $R^2$ indicates that approximately 13% of the total variance in the oral posttest performance of the two experimental groups can be explained by WM. Due to this significant role of WM in the learners’ posttest performance, a mixed ANCOVA was conducted on the oral performance of the two experimental groups with Time as a within-subjects variable (pretest, posttest, and delayed posttest), Group as a between-subjects variable (recast vs. metalinguistic), and WM as a covariate. The only significant difference observed was concerned with the Time variable, $F(1.531, 41.329) = 86.154, p < .05 (p = .000, \eta_p^2 = .761)$; the Greenhouse-Geisser correction was applied to the degrees of freedom due to the violation of sphericity. Three Bonferroni-corrected pairwise comparisons showed that overall, the experimental groups performed significantly better on the posttest than on the delayed posttest as well as on the pretest. Also, the groups showed significantly higher accuracy in the delayed posttest than in the pretest, reflecting the above-mentioned results regarding the performance of the recast group when the data from those who attended all three test sessions were analyzed (i.e., posttest > delayed posttest > pretest).
Learner performance on the posttest based on the data from those who showed up for the pretest and posttest sessions only was also regressed on WM and attention control measures. It was not significantly related to the cognitive variables (see Table I2 in Appendix I). Two sets of three separate regression analyses, one on pre-to-post gain scores and the other on pre-to-delayed gain scores, were also performed (see Table I3 in Appendix I). WM played a significant role in the pre-to-post development of the two experimental groups with a medium effect size, $F(1, 34) = 4.444, p < .05$ ($p = .042, R = .340, R^2 = .116, \text{adjusted } R^2 = .090$). WM accounted for approximately 9% of the total variance in the pre-to-post development of the two experimental groups.

Due to this significant relationship between WM and the pre-to-post development of the groups, an ANCOVA was conducted on the pre-to-post gains scores of the two experimental groups with their WM as a covariate. There was no significant difference between the recast and metalinguistic groups in their pre-to-post development, $F(1, 33) = .000, p = .998, \eta_p^2 = .000$. WM did not seem to change the overall results obtained in earlier analyses, that is, the two experimental groups were not significantly different from each other in their oral performance.

Lastly, to further explore the association between WM/attention control and the observed treatment effects (beneficial effects of recasts and metalinguistic feedback), the pre-to-post and pre-to-delayed gain scores of each experimental group were regressed on the two cognitive variables. As for the recast group (see Table I4 in Appendix I), the RT cost significantly predicted the pre-to-post gain scores, showing a somewhat large effect size, according to Cohen (1988), $F(1, 15) = 5.174, p < .05$ ($p = .038, R = .506, R^2 = .256, \text{adjusted } R^2 = .207$). The standardized regression coefficient, $\beta$, is -.506, which is the same as a correlation coefficient between the RT cost and the pre-to-post development of the group. The minus sign, of course,
indicates a negative correlation between the two. In other words, learners with higher attention control (and thus with a relatively low time cost for task switching) performed better than those with lower attention control (and thus with a relatively high time cost for task switching). The adjusted \( R^2 \) of .207 implies approximately 21% of the total variance can be explained by the RT cost. The error cost or WM was found to be unrelated to the pre-to-post development of the recast group. Moving on to the metalinguistic group (see Table I4 in Appendix I), WM measured via the OSPAN task appears to have significantly predicted the pre-to-post development of the group, showing quite a large effect size \( F (1, 15) = 8.974, \beta = .612, p < .05 \) (<.009, \( R^2 = .374, \) adjusted \( R^2 = .333 \)). The adjusted \( R^2 \) is .333, which means 33% of the total variance in the pre-to-post gain scores of the group can be accounted for by WM. The switch costs (both RT and error costs) were not significantly related to the pre-to-post development of the group.

Regarding pre-to-delayed gains scores, the only significant finding is that WMC was significantly predictive of the pre-to-delayed development scores of the metalinguistic group (see Table I5 in Appendix I), showing a large effect size, \( F (1, 13) = 5.407, \beta = .542, p < .05 \) (<.037, \( R^2 = .294, \) adjusted \( R^2 = .239 \)). The adjusted \( R^2 \) of .239 indicates that approximately 24% of the total variance in the pre-to-delayed development of the metalinguistic group can be accounted for by WM.

**Summary of Oral Performance on Past Unreal Conditional**

Both experimental groups performed significantly better than the control group on the two posttests and displayed significantly higher accuracy in the two posttests than in the pretest. No significant performance difference was found between the two experimental groups. No significant effect of Time was observed in the performance of the control group. Similarly, both experimental groups obtained significantly higher pre-to-post/pre-to-delayed gain scores than did
the control group. No significant difference was found between the two experimental groups in terms of gain scores. Regarding the relationship between WM/attention control and the observed benefits of corrective feedback, the RT cost significantly predicted the beneficial effects of recasts ($p < .05$). Approximately 21% of the total variance of the pre-to-post development of the recast group could be explained by the RT cost. Also, WMC was found to be associated with the pre-to-post development of the metalinguistic group ($p < .05$); 33% of the total variance of their pre-to-post gain scores could be accounted for by WMC. Furthermore, the pre-to-delayed development of the metalinguistic group was also significantly related to WMC ($p < .05$); 24% of the total variance of the development could be accounted for by WMC (Table 17 for a summary of learner performance on the past unreal conditional).
### Summary of Results on English Past Unreal Conditional

<table>
<thead>
<tr>
<th>All 3 Tests Involved</th>
<th>GJT Performance</th>
<th>Oral Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Posttest</td>
<td>Recast = Meta = Control</td>
<td>Recast, Meta &gt; Control, Recast = Meta</td>
</tr>
<tr>
<td>Delayed posttest</td>
<td>Recast = Meta = Control</td>
<td>Recast, Meta &gt; Control, Recast = Meta</td>
</tr>
<tr>
<td>Recast</td>
<td>Posttest, Delayed &gt; Pretest, Posttest = Delayed</td>
<td>Posttest &gt; Delayed &gt; Pretest</td>
</tr>
<tr>
<td>Metalinguistic</td>
<td>Posttest &gt; Pretest, Posttest = Delayed, Pretest = Delayed</td>
<td>Posttest, Delayed &gt; Pretest, Posttest = Delayed</td>
</tr>
<tr>
<td>Control</td>
<td>Pretest = Posttest = Delayed</td>
<td>Pretest = Posttest = Delayed</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Two Tests Involved</th>
<th>GJT Performance</th>
<th>Oral Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Posttest</td>
<td>Recast = Meta = Control</td>
<td>Recast, Meta &gt; Control, Recast = Meta</td>
</tr>
<tr>
<td>Recast</td>
<td>Posttest &gt; Pretest</td>
<td>Posttest &gt; Pretest</td>
</tr>
<tr>
<td>Metalinguistic</td>
<td>Posttest &gt; Pretest</td>
<td>Posttest &gt; Pretest</td>
</tr>
<tr>
<td>Control</td>
<td>Posttest &gt; Pretest</td>
<td>Posttest = Pretest</td>
</tr>
</tbody>
</table>

| Pre-to-Post Gain     | Recast > Control, Recast = Meta, Meta = Control | Recast, Meta > Control, Recast = Meta |
| Pre-to-Delayed Gain  | Recast = Meta = Control | Recast, Meta > Control, Recast = Meta |

<table>
<thead>
<tr>
<th>Cognitive Factors</th>
<th>GJT Performance</th>
<th>Oral Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>WMC</td>
<td>No significant relationship</td>
<td>Posttest(^a), Pre-to-post gain(^b), Pre-to-post gain (Meta), Pre-to-delayed gain (Meta)</td>
</tr>
<tr>
<td>RT Cost</td>
<td>No significant relationship</td>
<td>Pre-to-post gain (Recast)</td>
</tr>
<tr>
<td>Error Cost</td>
<td>No significant relationship</td>
<td>No significant relationship</td>
</tr>
</tbody>
</table>

**Note.** “=” means no statistically significant difference.

\(^a\)This finding is based on the results of the posttest in the pretest-and-posttest only dataset.

\(^b\)This finding is based on the pre-to-post gain scores of the two experimental groups as a whole.
Chapter V: Discussion

Research Questions and Overall Results

The dissertation research was designed, as an expanded version of Goo (in press), to investigate differential effects of recasts and metalinguistic feedback on the acquisition of the English that-trace filter and the past unreal conditional and whether/how WM and attention control mediate the extent to which L2 learners benefit from recasts and/or metalinguistic feedback.

With respect to the first research question, What are the relationships between the type of corrective feedback and the type of target structure in interaction-driven L2 learning? the results described above indicate a potential interaction between the type of corrective feedback and the type of structure. Metalinguistic feedback was more effective than recasts at facilitating accuracy in oral production involving the English that-trace filter, whereas recasts were slightly more effective than metalinguistic feedback at promoting the acquisition of grammatical knowledge on the English past unreal conditional (i.e., GJT performance). The metalinguistic group produced significantly more accurate sentences involving the that-trace filter, compared to the recast group as well as the control group, as shown in their posttest performance and pre-to-post gain scores. Note, however, that the relative advantage of metalinguistic feedback over recasts was observed only in oral production, but not in the GJT performance. Recasts, on the other hand, had a longer-lasting effect on the acquisition of knowledge on the English past unreal conditional compared to metalinguistic feedback, as evidenced in learner performance on the GJT. The recast group performed significantly better on the delayed posttest than on the pretest, whereas the metalinguistic group performed no better on the delayed posttest than on the pretest. The observed benefits of recasts lasted much longer than those of metalinguistic feedback. In addition,
it is only the recast group that obtained significantly higher pre-to-post gain scores than did the control group. No significant difference was found between the metalinguistic and control groups in terms of pre-to-post gain scores. No evidence of noticeable differences between the two feedback conditions was observed in oral production of the English past unreal conditional.

As for the second research question, *How is the effectiveness of corrective feedback at facilitating L2 learning mediated by individual differences in cognitive capacities?* the dissertation research provides partial evidence that individual differences in WM and attention control mediate the extent to which recasts and/or metalinguistic feedback benefit L2 learning. The evidence, however, was found only in the oral production of past unreal conditional sentences. Individual differences in WMC affected the extent to which metalinguistic feedback precipitated accuracy in producing past unreal conditional sentences; WM significantly predicted the pre-to-post and pre-to-delayed gain scores obtained by the metalinguistic group. Recasts (the benefits of recasts) were not found to be associated with WM differences. On the other hand, attention control measured via the task-switching paradigm was found to be predictive of the observed beneficial effects of recasts, but not of metalinguistic feedback, on the oral production of past unreal conditional sentences. No significant evidence was found of the involvement of WM or attention control in the acquisition of the *that*-trace filter. Discussed in what follows are potential contributing factors to, and specific issues related to, the findings in relation to the two research questions examined in this dissertation research.

**Discussion on Findings and Related Issues**

First and foremost, the differing effects of recasts and metalinguistic feedback evidenced in the present experiment resulted most likely from the different characteristics of the two target structures illustrated in an earlier chapter. The English *that*-trace filter is a purely syntactic
regulation that applies to *wh*-movement from the subject position of an embedded clause. No morphological change or semantic understanding is required. However, producing a grammatically accurate past unreal conditional sentence necessitates a clear understanding of the morphosyntactic constraint that entails changes to forms in both clauses and attention to meaning at the same time. That is, a more complex set of sub-rules is involved in the conversion of two clauses into a past unreal conditional sentence, compared to the that-trace filter that only requires the deletion of *that*. For this reason, the that-trace filter is somewhat simpler and thus easier to learn than the past unreal conditional through metalinguistic feedback, in that metalinguistic feedback provides rather direct and straightforward information on the that-trace filter by telling learners not to use the conjunction *that* when asking about the subject of the embedded/subordinate clause. Especially, metalinguistic feedback in the present study delivers much explicit information, compared to metalinguistic information operationalized in previous research, by specifically mentioning the grammatical term *conjunction* and what it denotes as in

*You are not supposed to use the conjunction *that* when you...* or *Don’t use the conjunction *that* when you...*

On the contrary, recasts in response to the violation of the that-trace filter does not seem salient enough to draw learners’ attention to the syntactic constraint and lead learners to produce a correct utterance that conforms to the that-trace filter because they involve reformulating learners’ utterances only by deleting the non-meaning-bearing sentential constituent *that*, which can be deleted in various instances, especially in informal speech or writing (e.g., at the beginning of clauses that follow such verbs as *say, think*, etc.). Furthermore, recasts operationalized in the present study were full/long recasts, the corrective nature of which may often go unnoticed, as opposed to partial/short recasts found to be more effective in drawing
learners’ attention to their linguistic gaps (see Egi, 2007b; Loewen & Philp, 2006; Philp, 2003; Sheen, 2006 for more discussion). It may have been quite challenging for some learners in the recast group to notice the corrective nature of recasts provided during interaction due to this low perceptibility or noticeability of the required change, the deletion of *that*.

Important to note is that the relative advantage of metalinguistic feedback over recasts in the acquisition of the *that*-trace filter was not observed in the GJT performance. No statistically significant difference between the two experimental groups was obtained in their GJT performance. Both groups significantly outperformed the control on their posttest and showed significantly higher pre-to-post gain scores than did the control group as shown in the analyses conducted on the data from the pretest and the posttest only. Given the fact that quite a few learners did not attend the delayed session, the relevant results (i.e., meta > control, meta = recast, and recast = control) based on the data from only those who participated in all three test sessions by no means provide any convincing evidence for the relative advantage of metalinguistic feedback over recasts. Thus, it is not unreasonable to contend that the advantage of metalinguistic feedback shown in the oral performance did not transfer to the GJT performance.

Contrary to the above-mentioned finding observed in the case of the *that*-trace filter in the oral production task as well as previous studies that compared recasts with metalinguistic feedback and reported the superiority of metalinguistic feedback over recasts (e.g., Ellis, 2007; Ellis et al., 2006; Sheen, 2007), recasts in the present study were found to be slightly more effective at precipitating the acquisition of knowledge on the past unreal conditional than metalinguistic feedback. The observed relative advantage of recasts over metalinguistic feedback may have been attributed to repeated exposure to target-like utterances provided through recasts. Unlike recasts in response to the violation of the *that*-trace filter, reformulated utterances and the
corrective nature of recasts may have been much more clearly delivered to the learners of the recast group due to the nature of the structure. The process of producing a past unreal conditional sentence inevitably involves directing learner attention almost at the entire sentence, which is quite long as they include two sentence-length clauses, by engaging learners to change the verbs both in the if clause and in the main/result clause. This focused attention necessary for the required changes to both clauses may have enabled the learners of the recast group to execute cognitive comparisons in a more effective and efficient way. In other words, recasts may have been provided in a more noticeable condition.

The metalinguistic group, however, was not exposed to targetlike sentences to compare with their own output, but just metalinguistic information including several linguistic terms (in L2) that they knew but had seldom used in an EFL context. As mentioned earlier, because producing a grammatically correct conditional sentence requires a somewhat drastic change to the verbs in both clauses in the sentence, some learners of the metalinguistic group may have needed target-like input to follow and/or compare with their own output. At least, opportunities for those unsure of the correctness of their output to test their hypotheses after receiving metalinguistic feedback would have helped them draw a clearer mental picture of the rule involving the past unreal conditional. However, as in Goo (in press), Leeman (2003), and Long et al. (1998), no modified output opportunities were allowed in the present study, unlike most feedback studies, where modified output opportunities were not controlled for (e.g., Ammar & Spada, 2006; Ellis, 2007; Ellis et al., 2006; Lyster, 2004; Lyster & Izquierdo, 2009; Sheen, 2007; Yang & Lyster, 2010); Lyster and Izquierdo (2009) blocked modified output opportunities for the recast condition, but allowed them for the prompt condition. Modified/pushed output is considered to be beneficial as it provides opportunities for learners to notice their linguistic
problems and test their hypotheses (Swain, 1985, 1995, 2005) and has been found to contribute
to L2 development (e.g., McDonough, 2005; Nobuyoshi & Ellis, 1993). Therefore, the learners
in the metalinguistic condition, by forfeiting modified output opportunities, may not have
received the entire benefits that metalinguistic feedback usually offers. Also, given recasts are
not as effective at eliciting modified output as prompts such as metalinguistic feedback (e.g.,
Lyster & Ranta, 1997; Panova & Lyster, 2002), and the effects of recasts may not depend upon
immediate production of modified output in any crucial way (e.g., Gass, 2003; Long, 2007;
Mackey & Philp, 1998), the blocking of modified output opportunities may have been much
more detrimental to the metalinguistic group, especially in learning such a complex structure as
the past unreal conditional, than to the recast group.

In brief, the blocking of modified output opportunities, the absence of target-like input,
and the provision of feedback in several linguistic terms that learners knew but had rarely used
may have affected the performance of some learners in the metalinguistic condition, leading to
the finding that metalinguistic feedback was not as effective as recasts at promoting the
acquisition of the past unreal conditional. Conversely, it may be contended that given its high
level of difficulty of the target, repeated exposure to target-like utterances (recasts) may have
played a nontrivial role in the learning of the past unreal conditional through recasts, contributing
much to the observed relative advantage of recasts over metalinguistic feedback in the present
experiment (see Leeman, 2003 for relevant discussion on positive evidence in recasts).

However, some findings on the performance of the control group on the GJT also deserve
explanations. Oddly, the control group significantly performed better on the posttest than on the
pretest ($p < .05, d = .58$) when the data from the pretest and the posttest only were considered,
and no significant difference was found among the groups. Unlike the that-trace filter, prior
learning likely played a role in the control group’s GJT performance. The structure is dealt with in most English grammar books designed for adult and adolescent EFL learners. Because the posttest was administered immediately after the interactional task on the same day, it is reasonably argued that the intensive interactional activity (Rashomon) rendered prior knowledge rekindled, no matter how complete the state of their prior knowledge on the target was, causing a positive impact on their posttest performance. For some learners in the control condition, the interactional task was then used as a practice opportunity. This, of course, is one obvious benefit of interaction (e.g., Mackey, 2007b; Swain, 1985, 1995, 2005). However, their improved performance by no means mitigates the overall findings of the beneficial effects of recasts and metalinguistic feedback on the acquisition of the past unreal conditional as clearly evidenced in the oral performance of the two experimental groups in comparison with that of the control group; see also the results of the GJT for more evidence.

Of particular interest are the findings observed in the present study in relation to the roles of WM and/or attention control in the beneficial effects of recasts and metalinguistic feedback on the acquisition of the target structures as well as the low correlation between WM and attention control. First of all, WM measured via OSPAN was not significantly correlated with attention/executive control measured through the task-set switching paradigm. Given the most critical underlying component of WM is attention/executive control, especially according to the executive attention theory (e.g., Engle, 2002; Engle, Tuholski, et al., 1999; Engle, Kane et al., 1999; Kane et al., 2007), no significant correlation between the two cognitive variables evidenced in the present study is quite a puzzling outcome. However, this unexpected finding has often been observed and reported in cognitive psychology (e.g., Colom, Shih, Flores-Mendoza, & Quiroga, 2006; Kane et al., 2007; Logan, 2004; Oberauer, Süß, Wilhelm, & Wittmann, 2003)
despite the postulated relationship between task-switching and executive control and relevant evidence (e.g., Baddeley, Chincotta, & Adlam, 2001; Liefooghe, Barrouillet, Vandierendonck, & Camos, 2008; Mayr & Kliegl, 2000, 2003; Rubinstein, Meyer, & Evans, 2001). Logan (2004), for instance, based on his results that showed a significant time cost of reconfiguring subordinate cognitive processes before performing a retrieved task in the task span procedure (Experiment 1), suggested that task switching requires cognitive processes that may operate above and beyond working memory that predominantly functions to changing goals or stimulus-response mapping rules in task switching. Similarly, Oberauer, Süß, Wilhelm, and Wittmann (2003) in their facet model of the structure of WM (two facets: content domains and cognitive functions) found that the task-set switching variables assumed to reflect executive functions (or supervision) did not substantially correlate with WMC factors (see their table, p. 182), posing a serious challenge to the executive attention theory (see Oberauer, Süß, Wilhelm, & Sander, 2007, for further discussion; see also Colom, et al., 2006, for a similar challenge to the executive attention theory). Kane et al. (2007) also reported their failed attempts to find an association between WMC and task switching. Nevertheless, in response to Oberauer et al.’s (2003, 2007) doubt regarding the primary role of cognitive control or executive attention processes in individual differences in WMC, Kane et al. (2007) argued that in most task-switching measures, learners are cued to switch between tasks, indicating that switch costs in this type of switching method may not be related to “volitional, executive-control processes” (p. 36). Because, as Logan (2004) suggested, there may exist multiple control processes involved in task switching, such a unitary model of WM as the executive attention theory (Engle, 2002; Engle, Kane, et al., 1999; Kane et al., 2007) may not explain the absence of evidence of the expected association between WM and task switching although evidence between WM and other simple attention control tasks has been
found in previous research (e.g., Conway et al., 2001; Kane et al., 2001; Kane & Engle, 2003). The exact nature of the relationship between WM and task switching is still inconclusive (see Vandierendonck et al., 2010 for discussion) despite the general consensus that task switching requires a cognitive processing ability in one way or another. The issue appears to merit much more research in the field of cognitive psychology given these controversial findings. Of course, it is beyond the scope of the present investigation.

With respect to the role of WM in L2 learning through corrective feedback, the present study provided evidence that WM may mediate beneficial effects of metalinguistic feedback, but not of recasts, on the acquisition of the past unreal conditional reflected in oral production. Three contributing factors to this finding are the type of structure, the design of task, and the type of dependent measure. Metalinguistic feedback includes several linguistic terms because producing a past unreal conditional sentence necessitates complex changes to the verbs in both clauses in a given sentence. That is, it was not a simple task to decode metalinguistic terms, process metalinguistic information without target-like input, and simply hold it in a readily accessible state for the next production opportunity. Additionally, as discussed above, modified output opportunities were blocked in the present study, and immediately after metalinguistic feedback, learners were prompted to move on to engage in producing a sentence involving the that-trace filter. Therefore, some learners may have had trouble recalling the metalinguistic information that they had received earlier when they needed to produce another past unreal conditional sentence in the following turn. In addition, learners may have felt psychological pressure in producing past unreal conditional sentences due to online cognitive demands. All three factors converged together and made it quite challenging for some learners, especially for those with relatively low WMC, to construct grammatically correct past unreal conditional sentences. On
the contrary, metalinguistic feedback on the *that*-trace filter used in the present study is much simpler than metalinguistic information on the past unreal conditional and clearly indicates that the conjunction “that” should be deleted. Thus, noticing may have occurred in a rather straightforward manner. In other words, cognitive control may not have been actively involved in the noticing of metalinguistic feedback, minimizing any potential role of WM.

Supportive evidence for this line of interpretation can be found in previous research on WMC in cognitive psychology, which indicates WM span differences arise only when a given task is cognitively challenging and thus requires cognitive control of attentional resources (e.g., Conway & Engle, 1994; Kane & Engle, 2000; Rosen & Engle, 1998). For instance, Conway and Engle (1994) showed that low-span participants were slower to respond to target items (letters and words) during verification tasks than high-span participants when each set item appeared as a target in two different memory sets, generating interference in the form of response competition (Experiments 1 and 2). The span differences were not such a crucial factor during their retrieval when each target item belonged to only one memory set (Experiments 3 and 4). That is, individual differences in WMC appear to play a crucial role in cognitively demanding tasks that require cognitive control and entail a competition for attentional resources. Kane and Engle (2000) also found that low-span participants did not differ from high-span participants in terms of accuracy in List 1 recall, that is, the first of three word lists (10 words in each list) from the same category (animals, occupations, or countries). It was evidenced, however, that low-span participants became more susceptible to PI than their high-span counterparts, as attested in their accuracy especially in List 3 recall. Similarly, Rosen and Engle (1998) observed that high- and low-span participants were much the same in the number of trials required to reach the learning criterion on the first list (i.e., without interference) in paired-associates tasks, but that high-span
participants were significantly less vulnerable to interference than were low-span participants in learning the second list under the interference condition; the same cue words used in the first list were also utilized in the second list but paired with different response words (e.g., bird-bath in List 1 and bird-dawn in List 2), which led to interference in the form of response competition. Given these findings in research on WM in the field of cognitive psychology, the present finding that WMC significantly predicted the beneficial effects of metalinguistic feedback on the acquisition of the past unreal conditional corresponds to previous empirical evidence in cognitive psychology that clearly indicates individual differences in WMC are most likely observable in attention-demanding tasks that require cognitive control.

With regard to WM in learning through recasts, given the present experiment was conducted in a dyadic laboratory setting as opposed to a classroom setting, for some learners, recasts involving the that-trace filter may have been understood with ease, but of course not for others. Noticing the absence of that in this lab-based setting is somewhat similar to an all-or-nothing phenomenon, that is, the learner either noticed it or did not notice it. Once it was noticed, then there was no need to provide any more recasts. If the learner was not aware of it, then it was likely that it was not noticed during the entire treatment. Thus, no middle ground seems to have existed in performance in this laboratory setting, which may have attenuated the role of cognitive control, resulting in no significant relationship between recasts and WMC/cognitive control. However, cognitive control seems to have played a mediating role in the efficacy of recasts on the learning of the past unreal conditional. Because, as discussed earlier, producing a past unreal conditional sentence requires focused attention due to the complexity of the structure, cognitive control emerges as a critical element in mental processing and determines the amount of attention available that can be assigned to such tasks as noticing recasts and cognitive
comparisons. In addition, there existed four different dimensions with respect to noticing: noticing the entire changes, noticing a change in the *if* clause, noticing a change in the main/result clause, and no noticing. This was similarly reflected in the scoring of their performance (i.e., 0, .5, and 1). In other words, their performance was diversified depending upon the level of attention that the learner could afford. From this viewpoint, it can be reasonably assumed that individual differences in cognitive or executive control may have affected the extent to which learners benefitted from recasts in learning the past unreal conditional, but not the *that*-trace filter.

However, there still remain three important unanswered questions: a) Why was cognitive control assumed to be involved in WMC unrelated to the benefits of recasts in learning the past unreal conditional? b) Why was cognitive control reflected in the task-set switching paradigm unrelated to the observed beneficial effects of metalinguistic feedback on the acquisition of the past unreal conditional? c) Why were individual differences in WMC and attention control unrelated to learner performance on the GJT on the past unreal conditional? As for the first two questions, it may be that the nature of different types of feedback interacts with the type of cognitive process. As Logan (2004) suggested, cognitive control reflected in task switching may require an extra process that is outside the boundary of WM. This extra process involves “reconfiguring the cognitive system” (Logan, 2004, p. 221) and occurs outside of WM which, according to Rubinstein et al. (2001), is involved in changing goals and rule-mappings, either during the overall process called task-set reconfiguration (Rogers & Monsell, 1995) or in the course of combating interference from previously activated rule-mappings and suppression (Allport et al., 1994; Allport & Wylie, 1999). Thus, this additional process that is not shared by WM and rather unique to cognitive control reflected in task switching may have led to this
difference. In the present experiment, recasts in learning the past unreal conditional may have evoked an additional cognitive process that enabled learners to compare recasts with their original nontargetlike utterances in a rather direct way during interaction due to their focused attention. This direct cognitive comparison is unlikely in the case of metalinguistic feedback due to its metalinguistic nature, the absence of targetlike input, and the blocking of modified output opportunities.

It is then speculated that individual differences in the ability to implement the above-mentioned additional cognitive process uniquely involved in cognitive comparisons may have mediated the extent to which learners benefitted from recasts as evidenced in the oral production task. However, given the present study found no significant role of cognitive control mirrored in task switching in the efficacy of recasts on the other target (that-trace filter), it appears to be a combination of the nature (and difficulty) of the target structure (past unreal conditional) and the functional nature of recasts that led to the observed relationship between recasts and cognitive control reflected in the task-set switching paradigm.

As for the third question, the nature of the GJT may have diluted any potential impact of the two cognitive factors as evidenced in the oral production task. As mentioned earlier, in the GJT, learners were given sufficient time for their judgment task on the past unreal conditional (15 seconds for each judgment item). In the GJT, learners were instructed to indicate whether each item was grammatically correct simply by putting a check mark on their answer sheet. This receptive and passive nature of the GJT as well as the provision of sufficient time for each judgment may also have caused relatively much less psychological pressure, compared to the amount of pressure that learners may have felt in the treatment task. These characteristics of the GJT administered in the present study are unlikely to reflect learners’ online cognitive demands.
involved in producing past unreal conditional sentences and processing corrective feedback
during interaction, rendering any mediating role of WM and/or attention control rather
unobservable. However, the oral production test was designed identically to the treatment task
(only with different contents in different drawings) and thus may have provided the same
cognitive conditions/demands as in the treatment task where cognitive control of attentional
resources was necessitated to produce past unreal conditional sentences. In other words, given its
incidental and time-pressured nature, the oral test likely imposed high attentional demands on
learners (e.g., Kormos, 2006) and made individual differences in WM and attention control more
visible. These characteristics of the two measures may have been attributed to the absence or
presence of relevant evidence.

Lastly, no significant relationship between recasts and WM with respect to the learning of
the English that-trace filter was evidenced in the present experiment, whereas Goo showed a
predictive role of WM in learning the same structure through recasts. It should be noted,
however, that the only comparable results are learner performance on the GJT because oral
production was not measured in Goo’s study, and written production was not measured in the
present study. This inconsistency between the two studies in terms of the GJT performance
appears to be attributable to the difference in the type of research setting. Goo’s study is a quasi-
experimental study conducted in a classroom setting where the corrective nature of recasts may
not always have been noticed by all learners in his study, whereas in the present study recasts
were provided in a dyadic laboratory setting where the noticing of recasts may have been more
or less straightforward for most of the learners in the recast group. As noted in Mackey and
Goo’s (2007) meta-analysis on interaction research, “differences between laboratory and
classroom settings may influence what learners perceive as important and that feedback may be
more likely to be perceived as such in laboratory settings, where it is presumably more consistent and focused” (p. 443); see also Nicholas et al. (2001) for similar discussion with respect to different findings on recasts. It is then assumed that the noticing of recasts may have occurred for many learners in the present laboratory experiment regardless of individual differences in cognitive control reflected in WMC. In other words, it is rather unlikely that individual differences in cognitive control of attentional resources led to any variability in learner performance in the present study. On the contrary, because much more focused attention may have been required to perform cognitive comparisons initiated by recasts in Goo’s classroom study, individual differences in cognitive/executive control mirrored in WMC may have resulted in the observed variability in the GJT performance of the recast group in Goo’s study despite the above-mentioned characteristics of the dependent variable measure that are not reflective of cognitive environments requiring cognitive control. A clearer picture of the role that WM may play in L2 learning through recasts may be obtained in future research including replications.

**Summary of Discussion**

First of all, the observed relative advantage of metalinguistic feedback over recasts in the oral production of the English *that*-trace filter appears to be a combined result of the nature of the target structure and the characteristics of the two feedback types provided during interaction. Metalinguistic feedback on the *that*-trace filter provided quite simple and straightforward information on the grammaticality of the learner’s oral production involving the *that*-trace filter, whereas the corrective nature of recasts on the *that*-trace filter may not have been delivered in a clear-cut manner due to the low perceptibility of the obligatory deletion of the non-meaning-bearing constituent, *that*, which can also be omitted in various other instances.
The study also found that recasts were slightly more effective at learning the past unreal conditional than metalinguistic feedback. Similarly to the case of metalinguistic feedback on the *that*-trace filter, the nature of the structure and the characteristics of the two feedback types likely contributed to the observed finding. Because producing a past unreal conditional sentence involves multiple changes and thus entails focused attention required to enable learners to compare their nontarlike utterances with targetlike reformulated versions, the likelihood of detecting recasts and noticing their corrective nature may have increased to a great extent. This heightened possibility of noticing recasts may have resulted in effective cognitive comparisons through repeated exposure to target-like utterances provided in recasts.

For some learners in the metalinguistic group, however, three potential factors may have delimited the beneficial effects of metalinguistic feedback on the acquisition of the past unreal conditional. Firstly, unlike recasts, no target-like input (e.g., through reformulations as in recasts) was provided for the learners of the metalinguistic group to compare with their non-target-like utterances. Secondly, modified output opportunities were blocked, that is, the learners forfeited one critical route to confirming their understanding by testing their own hypotheses on the past unreal conditional. Lastly, metalinguistic feedback on the past unreal conditional included several linguistic terms in L2 that they knew, but had rarely used, making it a more difficult task for some learners to understand the target structure. All three factors may have contributed, in combination, to the finding that metalinguistic feedback was not as effective as recasts at facilitating the acquisition of the past unreal conditional. The finding also suggests that repeated exposure to target-like utterances included in recasts may play an important role in learning such a complex structure as the past unreal conditional. However, due to the innate difficulty of the structure, the relative advantage of recasts over metalinguistic feedback was evidenced only in
the GJT performance. This indicates that producing past unreal conditional sentences in an incidental setting was still challenging for some learners in the recast group because it requires learners to focus on multiple aspects almost simultaneously. As Long (2007) noted, the overall findings with respect to the relationship between the type of structure and the type of feedback suggest that the more complex a target structure is, the more beneficial/effective recasts may be as compared to metalinguistic feedback. The improved performance of the control group seems attributable to the opportunity to test hypotheses on the target by participating in the interactional task, which may also have helped refresh their prior learning on the target, although no feedback was provided. However, the level of the improvement achieved by the control group is by no means as comparable as that of the experimental groups.

The present findings with regard to cognitive capacities are somewhat surprising. The expected relationship between WM and attention control was not observed. However, this seemingly deviant finding has also been observed in previous research in cognitive psychology (e.g., Kane et al., 2007; Logan, 2004; Oberauer et al, 2003). The potential culprit for this nonsignificant correlation between the two is, according to Logan (2004), for instance, an additional process (i.e., the reconfiguration of the cognitive system) that is required in the task-set switching paradigm, which is above and beyond the boundary of WM. The additional process seems to be at the core of the nonsignificant correlation between the two cognitive factors (see also Colom et al., 2006; Kane et al., 2007; Oberauer, 2003 for similar findings and relevant discussion).

As for individual differences in WMC, all three factors mentioned above with respect to metalinguistic feedback in relation to the relative advantage of recasts over metalinguistic feedback in learning the past unreal conditional may have contributed to the role of WM in the
oral performance of the metalinguistic group: the need to decode and maintain metalinguistic information including several linguistic terms, no modified output opportunity, and psychological pressure in oral production. These factors likely necessitated cognitive control of attentional resources in learning the past unreal conditional through corrective feedback. However, metalinguistic feedback on the that-trace filter was more or less straightforward and thus may not have required too much of cognitive control, resulting in no significant impact of individual differences in WM on the learning of the that-trace filter. Given WM differences arise only when a task is cognitively challenging (e.g., Conway & Engle, 1994; Kane & Engle, 2000; Rosen & Engle, 1998), it is rather unsurprising that WM was found to be significantly related to the effectiveness of metalinguistic feedback in learning the past unreal conditional, but not in learning the that-trace filter.

Cognitive control also played an important role in the efficacy of recasts on the acquisition of the past unreal conditional as evidenced in the oral production test. However, the effectiveness of recasts was mediated by cognitive control mirrored in the task switching paradigm, but not in WM. As Logan (2004) indicated, it may be due to an extra process (i.e., the reconfiguration of the cognitive system) involved in task switching. In terms of importance, this extra process is analogous to an additional process required to fully benefit from recasts, as compared to metalinguistic feedback, that is, a cognitive comparison between the learner’s original nontargetlike utterance and its targetlike reformulated version provided in a recast. This requirement of cognitive comparisons may have differentiated cognitive control reflected in task switching from cognitive/executive control in WM. Again, given psychological pressure the learners may have felt in oral production, it is not quite surprising that the impact of cognitive
control using the task-set switching paradigm was observed in learner performance on the oral test, but not on the GJT.

Interestingly, no significant role of WM and attention was found in the case of the *that*-trace filter. This nonsignificant finding is attributable to a combination of the type of research setting, the type of structure, and the scoring system in relation to the nature of the structure. Recasts were provided in a dyadic lab-based setting, and thus the deletion of the conjunction *that* may have been somewhat easy to detect for many of the learners in the corresponding group, albeit not for all. More importantly, since the only requirement is to delete *that*, their performance was scored in a dichotic manner, that is, it was all-or-nothing for each individual item. In other words, the distribution of learner performance provided rather a limited space for learner variability, whereas learner performance as to the past unreal conditional was scored in consideration of four possibilities (correct in both clauses, correct only in the if clause, correct only in the result clause, and incorrect in either clause), leading to much variability in learner performance.

When compared to Goo’s (in press) study that showed a significant mediating effect of WM on the extent to which learners benefitted from recasts in learning the *that*-trace filter, the present study failed to confirm his finding. However, this inconsistency may be attributable to two different research settings: classroom vs. laboratory settings. That is, recasts were relatively easier to notice in the present study conducted in a laboratory setting than in Goo’s quasi-experimental study conducted in a classroom setting. Therefore, statistical variability in learner performance in the present study seems much more limited than in Goo’s study, contributing to the observed nonsignificant role of WM.
Chapter VI: Limitations and Future Research

Notwithstanding the new and interesting insights into research on corrective feedback that the dissertation study may offer, it is still subject to some limitations. For instance, although each significant finding with respect to the association between corrective feedback and WM/attention control was based on a decent sample size ($n = 17$ for both significant findings, i.e. one between WM and metalinguistic feedback and the other between cognitive control and recasts), larger sample sizes would have resulted in statistically more robust outcomes or even different findings and interpretations. Similarly, it appears crucial to maintain original sample sizes throughout an experiment because changes in sample sizes may entail different statistical results and interpretations. For instance, the subject mortality observed in the present research led to different outcomes from two separate analyses (one on the data from the pretest and the posttest only and the other on the data from all three test sessions) and thus allowed somewhat different interpretations of the GJT performance of the recast group on the *that*-trace filter. More caution and efforts to prevent the subject mortality, albeit inescapable, would have contributed to furthering the internal validity of the findings and interpretations.

Also, given that the present research was conducted in a foreign language context, in which acquiring linguistic forms is usually prioritized as compared to meaningful communication, this contextual factor may have led the learners to pay more attention to linguistic accuracy than to meaning during the treatment activity. Mackey and Goo (2007) showed in their meta-analysis that interactional treatments had been more apparent in studies conducted in a foreign language context than those conducted in a second language context. In addition, interaction research in a laboratory setting, according to Mackey and Goo’s meta-analysis, showed larger effect sizes than in a classroom setting. That is, considering the present
experiment was lab-based and conducted in a foreign language context, the observed treatment effects may not be too surprising. In other words, the generalizability of the findings obtained here is rather limited to the context and setting involved in the present study. Other research contexts (e.g., SL context) and settings (e.g., classroom setting) may induce differing sets of outcomes. Also, because only Korean EFL learners participated in this laboratory experiment, it is not guaranteed that similar results to those observed in the present experiment would be obtained with EFL learners of other L1 backgrounds. The extent of treatment effects may vary depending upon the level of L1 influences that may be determined to some extent by L1-L2 similarities and differences. The impacts of these factors on research findings are nontrivial, as evidenced in Mackey and Goo’s meta-analysis, and thus they should be considered in future research.

Regarding the design of the interactional task (Rashomon), it was a tightly structured activity in which the learners were obliged to produce sentences involving the two targets. This task-essentialness in Loschky and Bley-Vroman’s (1993) sense likely heightened the intensity of treatments so that the learners may have focused their attention on the target structures/rules, and hence corrective feedback may have become salient for quite a few learners. Given this type of intensive treatment is rather unlikely in most learning settings, it should be admitted that the so-called ecological validity is, obviously, not high enough to generalize the findings of the present dissertation research to all learning settings/contexts. Although this is one characteristic of most laboratory research, it is worth trying to investigate whether similar results can be obtained with less structured treatment activities that may be utilized in classroom as well as research settings.

As discussed earlier, previous interaction research has shown that the type and nature of a target structure may interact with the effectiveness of corrective feedback, affecting the extent to
which learners benefit from particular feedback (e.g., Ellis, 2007; Iwashita, 2003; Leeman, 2003; Long et al., 1998; Sagarra, 2007). The present dissertation study provides another important piece of evidence for this claim by showing recasts and metalinguistic feedback differed in their impacts on the acquisition of the two target structures. However, other linguistic features in English (phonological features as well as syntactic and morphosyntactic aspects of English) or features in other languages also need to be tested in order to further verify the claim and to diversify our insights into L2 learning phenomena via interaction. This line of research may also make significant contributions to WM research. Sagarra and Herschensohn (2010) found that higher-WMC intermediate learners were more sensitive to gender agreement violations than lower-WMC intermediate learners, whereas WM was not correlated with number agreement violations. This finding indicates that the predictive role of WM in L2 processing depends on the type of target, which is also evidenced in the present dissertation research. This potential interaction between WM and different target forms merits further research. Additionally, replications of the present research as well as of Goo’s (in press) study should also be considered in future research as they likely contribute to deepening our understanding by confirming or disconfirming the relevant findings obtained in the two studies.

Also, although this dissertation study was not designed to investigate the role of proficiency in L2 learning through corrective feedback, it definitely merits future research given some interaction research has identified learner proficiency as a potential mediating factor that may affect the effectiveness of corrective feedback (e.g., Ammar & Spada, 2006; Trofimovich et al., 2007). It should be noted, however, that although proficiency and developmental readiness are different, they are closely related to each other. Therefore, when proficiency is examined with regard to its relation to L2 learning in an empirical study, it should be considered whether
L2 learners are developmentally ready to learn a particular target structure. Put differently, if L2 participants are not developmentally ready, for instance, the amount of reasonable information that can be obtained as to the role of proficiency in L2 learning through corrective feedback is highly limited. In this regard, operationalizing proficiency in consideration of developmental readiness is a critical task for researchers interested in how L2 proficiency mediates the extent of feedback effects. Nevertheless, there is no denying that proficiency deserves more research in terms of its relation to the extent of the effectiveness of corrective feedback. The issue of L2 proficiency is also an interesting area of inquiry in future research because it is quite conceivable that L2 proficiency may interact with individual differences in WM and/or attention control in L2 learning contexts. In fact, Walter (2004) showed that individual differences in WMC was significantly correlated with the performance of low-intermediate learners on L2 reading comprehension, but no significant correlation was observed between WM and the performance of high-intermediate learners on L2 reading comprehension. More research on the role of proficiency in L2 learning in relation to the type of structure, the type of corrective feedback, and cognitive capacities such as WM and attention control will surely contribute to broadening our knowledge structure regarding the potential impacts of these factors on L2 learning.

Another important methodological issue to consider in future research is related to the use of other WM span tasks and attention control measures. Although an OSPAN task has been one of the most frequently used tasks along with a reading span task to tap verbal WM and proven to be a valid measure (see Conway et al., 2005 for a review), it is, of course, not “the gold standard [italics in original]” (Conway et al., 2005, p. 784) measure of WMC. Therefore, it is desirable to employ multiple WM span measures so that more valid and reliable results can be obtained regarding L2 learners’ WM. Also, the OSPAN task used in the present study is a more
or less language-independent measure, but utilizing L1 and L2 WM span tasks together in one study (e.g., Geva & Ryan, 1993; Juffs, 2004, 2005; Miyake & Friedman, 1998; Mackey et al., 2002; Walter, 2004) may provide diverse insights into the relationship between WM and L2 learning through corrective feedback insofar as L2 WM measures are appropriate for proficiency. Despite high correlations between L1 and L2 for high-proficient L2 learners (e.g., Osaka & Osaka, 1992; Osaka, Osaka, & Groner, 1993), it is still a worthwhile attempt to explore how L1 and L2 WM span measures are related to the effectiveness of corrective feedback on L2 learning because being highly correlated is not isomorphic with being identical.

With regard to attention control measures, considering the finding that WM was not correlated with attention control measured via a task-set switching test in the present research, other attention control measures that have successfully shown the relationship between WM and attention control such as Stroop tasks (Kane & Engle, 2003) and antisaccade tasks (Kane et al., 2001) need to be employed in future research to explore any potential association among corrective feedback, WM, and attention control. Also, measures to tap STM (e.g., word span, letter span, digit span) may provide additional information with respect to the role of cognitive capacities in L2 learning through corrective feedback. In general, WM is a better predictor of performance on complex cognitive tasks than STM (Conway et al., 2002; Daneman & Merikle, 1996; Engle, Tuholski et al., 1999). For instance, Conway et al. (2002) and Engle, Tuholski et al. (1999) in their latent variable analyses showed that the residual WMC variance (after the common component to WMC and STM is statistically factored out) was more strongly correlated with general fluid intelligence than was the residual STM variance (after the common component to WMC and STM is factored out). However, given WM and STM are strongly correlated with each other (e.g., Engle, Kane, et al., 1999; Engle, Tuholski, et al., 1999; Kane, 2007), STM may
be linked, albeit in a somewhat different way, to the efficacy of corrective feedback on L2 learning. Relevant findings in future interaction research involving STM may add another interesting layer to our knowledge on whether and how cognitive capacities mediate the effectiveness of corrective feedback in L2 learning.

Another issue that can be examined in future research is whether and how individual differences in cognitive capacities (WM, STM, and attention control) are associated with the effectiveness of other forms of implicit and explicit corrective feedback in L2 learning (e.g., clarification requests, explicit correction, elicitations, etc.). Because different forms of corrective feedback involve different levels of explicitness/implicitness (Doughty & Williams, 1998; Loewen & Nabei, 2007), it is likely that they entail different amounts of attention, and that cognitive control of attentional resources required to benefit from them may vary depending upon the type of feedback or even the level of explicitness within the same feedback type. For instance, short/partial recasts (as opposed to long recasts used in the present study), which are assumed to be easier to notice than long/full recasts (see Egi, 2007b; Loewen & Philp, 2006; Philp, 2003; Sheen, 2006 for more on this issue), may not require learners to exercise high-level cognitive control, and thus likely offer a different array of observations. Also, given phonological enhancement through emphatic stress makes recasts more explicit as opposed to implicit recasts, as Chaudron (1977) pointed out, and thus more noticeable, comparing more explicit recasts with less explicit recasts at suprasegmental levels in terms of a possible association between beneficial effects of recasts and individual differences in cognitive capacities may offer another important aspect of the role cognitive capacities play in L2 learning through corrective feedback. Therefore, the relationship between the level of explicitness within
the same feedback type (e.g., recasts) and individual differences in cognitive capacities is another attention-drawing research issue that should also be examined in the near future.

Lastly, due to the sizes of the samples for the treatment groups, extreme-groups comparisons in relation to the type of corrective feedback (e.g., recasts vs. metalinguistic feedback), the level of WMC (e.g., top 25% as high WM vs. bottom 25% as low WM), and the level of attention control (e.g., top 25% vs. bottom 25%) were not conducted in the present dissertation research. These extreme-groups comparisons, if feasible in future research, may offer additional insights into the potential implications of WM and/or attention control in the relative efficacy of one feedback type over another on L2 learning, which correlational analyses do not provide (e.g., more clear-cut and straightforward information on high WM vs. low WM learners, albeit not a full landscape due to the dropping of mid 50%).

In sum, despite the interesting findings on the relationship between corrective feedback and individual differences in cognitive capacities, the present dissertation study suffers several limitations affecting the generalizability of the findings, but offers various future research topics as well as methodological issues to be considered in future research (e.g., subject mortality, research setting/context, measures of other cognitive variables, treatment activities, dependent variable measures, proficiency, level of explicitness, and extreme-groups designs). More rigorous research that considers these methodological issues on potential research areas mentioned in this section may assure a more definitive picture of the relationship between corrective feedback and individual differences in cognitive capacities.
Chapter VII: Conclusion

The present study was designed to explore differential effects of recasts and metalinguistic feedback on the acquisition of two target structures: the English *that*-trace filter and the past unreal conditional. In addition, the study investigated whether and how WM and attention control mediate the extent to which learners benefit from these two feedback types. The OSPAN was utilized to measure WMC, and the task-set switching paradigm was employed to measure attention control. The results of the GJT's and the oral production tasks showed that metalinguistic feedback was more effective than recasts at facilitating the acquisition of the English *that*-trace filter. However, this finding was evidenced in the oral production task, but not in the GJT performance. The study also found that recasts were slightly more beneficial than metalinguistic feedback in the learning of the past unreal conditional. This relative effect of recasts over metalinguistic feedback was observed only in the GJT performance, but not in the oral production task. Considered together, the findings suggest that recasts may work more effectively with a relative more complex target structure and metalinguistic feedback with a less complex target structure.

As for the role of WM/attention control, the study showed that WM significantly predicted the pre-to-post development of the metalinguistic group in the oral production task on the past unreal conditional, whereas attention control reflected in task switching failed to predict the improved performance of the group on any measure. Interestingly, attention control significantly predicted the pre-to-post development of the recast group in the oral production task on the past unreal conditional, but WM was not found to mediate the improved performance of the recast group on any assessment measure. No significant association between WM/attention
control and the treatment effects was evidenced with regard to the acquisition of the English that-trace filter.

The relative effects of metalinguistic feedback over recasts in the acquisition of the that-trace filter may be attributed to the nature of feedback. Metalinguistic feedback is rather simple and straightforward as compared to recasts. Thus, the learners in the metalinguistic group somewhat easily noticed and understood metalinguistic feedback, whereas detecting the deletion of the conjunction that may not be an easy task for some learners in the recast group. In other words, the nature of feedback and the type of target seems to have contributed to this relative advantage of metalinguistic feedback over recasts in learning the that-trace filter. However, the recast group performed slightly better than the metalinguistic group in the acquisition of the past unreal conditional. Because learning the past unreal conditional involves multiple changes, it requires focused attention, and this high level of attention may have enabled learners to detect a target-like reformulated version provided via a recast, which let them to engage in a cognitive comparison. However, the learners of the metalinguistic group may have had difficulty in decoding metalinguistic information including several linguistic terms and maintaining the information until the next production turn. The blocking of modified output opportunities may also have affected their understanding of the past unreal conditional by removing their opportunities to confirm their understanding. These factors may have been attributed to the relative advantage of recasts observed in the present study.

Interestingly, no significant relationship was found between WM and attention control. This may be because cognitive control measured via the task-set switching paradigm may involve an extra process that is unique to attention control in task switching but outside of WM. This extra process may have differentiated the role of cognitive control measured via task
switching from the role of WM measured via the OSPAN task in the effectiveness of recasts and metalinguistic feedback in the acquisition of the past unreal conditional. Considering the difficulty of processing and maintaining metalinguistic information without target-like input and modified output opportunities, the result that WM was found to be related to metalinguistic feedback confirms the previous finding that suggests individual differences in WMC emerge only during an attention-deman
ding task. However, because cognitive control in task switching is assumed to involve an extra cognitive process, no significant relationship was evidenced between attention control and the oral performance of the metalinguistic group. On the contrary, a significant relationship was observed between attention control and the improved performance of the recast group on the oral task, but WM did not play a role in the oral performance of the group. The scoring methods may have contributed to no significant association between WM/attention control and learner performance on either measure regarding the that-trace filter. The inconsistency observed in terms of the relationship between WM and recasts may have been attributed to the type of research setting. That is, the present study was conducted in a laboratory setting where recasts may have been much more clearly delivered compared to a classroom setting in which Goo’s study was carried out. Noticing recasts in a classroom setting may have been a challenging task so that cognitive control in WM may have played a mediating role, which is not the case in the present study.

As discussed in the previous section, several issues appear to merit future research: proficiency, contexts/settings, other structures, other measures of cognitive capacities, other forms of feedback (or the same type of feedback with different levels of explicitness), and extreme-groups comparisons. Future research on (or in consideration of) these factors is likely to bring to light more insights into any potential mediating roles of cognitive capacities in the
efficacy of corrective feedback on L2 learning, providing more convincing evidence for, or evidence against, the present findings. Likewise, it is of particular importance to conduct similar studies to, or replications of, the present study as well as Goo’s (in press) study so that the observed findings in both studies can be confirmed or disconfirmed in our concerted efforts to obtain a clearer picture of the relationship between corrective feedback and cognitive capacities.
Notes

1 General fluid intelligence \((gF)\) refers to “the ability to solve novel problems and adapt to new situations and is thought to be nonverbal and relatively culture free” (Engle, Tuholski, et al., 1999, p. 313), which is independent of general knowledge. Two standardized tests most frequently used to measure \(gF\) are Cattell’s Culture Fair Test and Raven’s Standard Progressive Matrices.

2 Given means and standard deviations, statistical significance might have been found between the pretest and the immediate posttest and between the pretest and the delayed posttest if mixed ANOVAs had been used.

3 Again, as they did not carry out a mixed or repeated-measures ANOVA, no significance test value was reported, but the differences between the pretest and the immediate posttest and between the pretest and the delayed posttest would have reached statistical significance if a mixed ANOVA had been conducted.

4 The dichotic listening task/procedure is to repeat aloud the message (or words) presented to one ear while ignoring information presented to the other ear.

5 In the Stroop task, participants are required to name each color word presented on the screen based on the color of the ink in which the word is printed; for instance, if \(RED\) is presented in blue color, they are supposed to say “Blue” not “Red.” Kane and Engle (2003) showed that low-span subjects were more susceptible to Stroop interference effects than high-span subjects, committing more color-naming errors than did their high-span counterparts, especially in the 75% or 80% congruent conditions.

6 In both prosaccade and antisaccade tasks, participants fixate in the middle of a visual display but must respond to each target stimulus (preceded by an attention-attracting cue)
presented randomly to the left or right side of the display. Whereas the attention-attracting cue and the target stimulus appear on the same side of the display in the prosaccade task, the cue always appears on the opposite side of the display from the target in the antisaccade task (e.g., a cue on the left side of fixation followed by a target stimulus on the right side).

Proactive interference is defined as “the generally disruptive effect of prior learning on the ability to retrieve more recently learned information” (Lustig & Hasher, 2002, p. 90) assumed to build up in the traditional ascending order of item presentation (see Lustig, May, & Hasher, 2001; May, Hasher, & Kane, 1999 for research on this ordering issue, ascending vs. descending). In the traditional ascending order of presentation, large set sizes crucial for high recall scores appear later, and prior trials likely cause PI when large sets of items need to be recalled, negatively affecting overall performance on WM span tasks. This issue of PI is taken into consideration.
Appendix A

Oral Production Sheet for Learners

Type A

1. Bank Robbery

Young-soo thinks that _____________ robbed the bank yesterday.

Dong-hyun thinks that _____________ robbed the bank yesterday.

Cheol-soo thinks that _____________ robbed the bank yesterday.

Joon-ho thinks that _____________ robbed the bank yesterday.

Kwang-mo thinks that _____________ robbed the bank yesterday.

Sang-goo thinks that _____________ robbed the bank yesterday.

2. Going to the movies with Sue

Sang-goo thinks that _____________ went to the movies with Sue last night.

Young-soo thinks that _____________ went to the movies with Sue last night.

Joon-ho thinks that _____________ went to the movies with Sue last night.

Dong-hyun thinks that _____________ went to the movies with Sue last night.

Cheol-soo thinks that _____________ went to the movies with Sue last night.

Kwang-mo thinks that _____________ went to the movies with Sue last night.

Treatment

1. Stealing watches from a jeweler’s

Cheol-soo thinks that _____________ stole them from the jeweler’s last week.

Young-soo thinks that _____________ stole them from the jeweler’s last week.
Joon-ho thinks that ____________ stole them from the jeweler’s last week.

Kwang-mo thinks that ____________ stole them from the jeweler’s last week.

Sang-goo thinks that ____________ stole them from the jeweler’s last week.

Dong-hyun thinks that ____________ stole them from the jeweler’s last week.

2. Breaking into Mary’s house.

Young-soo thinks that ____________ broke into Mary’s house last night.

Dong-hyun thinks that ____________ broke into Mary’s house last night.

Sang-goo thinks that ____________ broke into Mary’s house last night.

Cheol-soo thinks that ____________ broke into Mary’s house last night.

Joon-ho thinks that ____________ broke into Mary’s house last night.

Kwang-mo thinks that ____________ broke into Mary’s house last night.

3. Having cancer

Kwang-mo thinks that ____________ has cancer.

Young-soo thinks that ____________ has cancer.

Dong-hyun thinks that ____________ has cancer.

Joon-ho thinks that ____________ has cancer.

Cheol-soo thinks that ____________ has cancer.

Sang-goo thinks that ____________ has cancer.

Type B

1. Drinking Jay’s coke containing a cigarette butt.
Young-soo thinks that ______________ drank Jay’s coke.
Dong-hyun thinks that ______________ drank Jay’s coke.
Cheol-soo thinks that ______________ drank Jay’s coke.
Joon-ho thinks that ______________ drank Jay’s coke.
Kwang-mo thinks that ______________ drank Jay’s coke.
Sang-goo thinks that ______________ drank Jay’s coke.

2. Smoking in a movie theater.
Dong-hyun thinks that ______________ smoked in a movie theater.
Cheol-soo thinks that ______________ smoked in a movie theater.
Kwang-mo thinks that ______________ smoked in a movie theater.
Sang-goo thinks that ______________ smoked in a movie theater.
Young-soo thinks that ______________ smoked in a movie theater.
Joon-ho thinks that ______________ smoked in a movie theater.

Type C
1. Smashing a laptop computer.
Sang-goo thinks that ______________ smashed Jeff’s laptop computer.
Young-soo thinks that ______________ smashed Jeff’s laptop computer.
Dong-hyun thinks that ______________ smashed Jeff’s laptop computer.
Cheol-soo thinks that ______________ smashed Jeff’s laptop computer.
Joon-ho thinks that ______________ smashed Jeff’s laptop computer.
Kwang-mo thinks that ______________ smashed Jeff’s laptop computer.
2. Murdering a pizza delivery guy

Cheol-soo thinks that ____________ murdered the delivery guy.

Joon-ho thinks that ____________ murdered the delivery guy.

Kwang-mo thinks that ____________ murdered the delivery guy.

Sang-goo thinks that ____________ murdered the delivery guy.

Young-soo thinks that ____________ murdered the delivery guy.

Dong-hyun thinks that ____________ murdered the delivery guy.
## Appendix B

**OSPAN Math**

School: ___________________________  Name: ____________________________

### Practice

<table>
<thead>
<tr>
<th></th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>26</td>
<td></td>
<td></td>
</tr>
<tr>
<td>27</td>
<td></td>
<td></td>
</tr>
<tr>
<td>28</td>
<td></td>
<td></td>
</tr>
<tr>
<td>29</td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td></td>
<td></td>
</tr>
<tr>
<td>31</td>
<td></td>
<td></td>
</tr>
<tr>
<td>32</td>
<td></td>
<td></td>
</tr>
<tr>
<td>33</td>
<td></td>
<td></td>
</tr>
<tr>
<td>34</td>
<td></td>
<td></td>
</tr>
<tr>
<td>35</td>
<td></td>
<td></td>
</tr>
<tr>
<td>36</td>
<td></td>
<td></td>
</tr>
<tr>
<td>37</td>
<td></td>
<td></td>
</tr>
<tr>
<td>38</td>
<td></td>
<td></td>
</tr>
<tr>
<td>39</td>
<td></td>
<td></td>
</tr>
<tr>
<td>40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>41</td>
<td></td>
<td></td>
</tr>
<tr>
<td>42</td>
<td></td>
<td></td>
</tr>
<tr>
<td>43</td>
<td></td>
<td></td>
</tr>
<tr>
<td>44</td>
<td></td>
<td></td>
</tr>
<tr>
<td>45</td>
<td></td>
<td></td>
</tr>
<tr>
<td>46</td>
<td></td>
<td></td>
</tr>
<tr>
<td>47</td>
<td></td>
<td></td>
</tr>
<tr>
<td>48</td>
<td></td>
<td></td>
</tr>
<tr>
<td>49</td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td></td>
<td></td>
</tr>
<tr>
<td>51</td>
<td></td>
<td></td>
</tr>
<tr>
<td>52</td>
<td></td>
<td></td>
</tr>
<tr>
<td>53</td>
<td></td>
<td></td>
</tr>
<tr>
<td>54</td>
<td></td>
<td></td>
</tr>
<tr>
<td>55</td>
<td></td>
<td></td>
</tr>
<tr>
<td>56</td>
<td></td>
<td></td>
</tr>
<tr>
<td>57</td>
<td></td>
<td></td>
</tr>
<tr>
<td>58</td>
<td></td>
<td></td>
</tr>
<tr>
<td>59</td>
<td></td>
<td></td>
</tr>
<tr>
<td>60</td>
<td></td>
<td></td>
</tr>
<tr>
<td>61</td>
<td></td>
<td></td>
</tr>
<tr>
<td>62</td>
<td></td>
<td></td>
</tr>
<tr>
<td>63</td>
<td></td>
<td></td>
</tr>
<tr>
<td>64</td>
<td></td>
<td></td>
</tr>
<tr>
<td>65</td>
<td></td>
<td></td>
</tr>
<tr>
<td>66</td>
<td></td>
<td></td>
</tr>
<tr>
<td>67</td>
<td></td>
<td></td>
</tr>
<tr>
<td>68</td>
<td></td>
<td></td>
</tr>
<tr>
<td>69</td>
<td></td>
<td></td>
</tr>
<tr>
<td>70</td>
<td></td>
<td></td>
</tr>
<tr>
<td>71</td>
<td></td>
<td></td>
</tr>
<tr>
<td>72</td>
<td></td>
<td></td>
</tr>
<tr>
<td>73</td>
<td></td>
<td></td>
</tr>
<tr>
<td>74</td>
<td></td>
<td></td>
</tr>
<tr>
<td>75</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Operation Span

Name_________________

****************************
Practice

a) ___________     ____________
b) ___________     ____________
c) ___________     ____________

*****************************

1)    ________     ________     ________     ________     ________     ________     ________

2)    ______     ________     ________     ________     ________     ________     ________

3)    ________     ________     ________     ________     ________     ________     ________

4)    ________     ________     ________     ________     ________     ________     ________

5)    ________     ________     ________     ________     ________     ________     ________

6)    ________     ________     ________     ________     ________     ________     ________

7)    ________     ________     ________     ________     ________     ________     ________

8)    ________     ________     ________     ________     ________     ________     ________

9)    ________     ________     ________     ________     ________     ________     ________

10)  ________     ________     ________     ________     ________     ________     ________

11)  ________     ________     ________     ________     ________     ________     ________

12)  ________     ________     ________     ________     ________     ________     ________

13)  ________     ________     ________     ________     ________     ________     ________

14)  ________     ________     ________     ________     ________     ________     ________

15)  ________     ________     ________     ________     ________     ________     ________
Appendix C

Grammaticality Judgment Test Items (That-trace filter)

Type A

# 1.
They think that Sarah developed a rocket last year.
What do they think that Sarah developed last year?

# 2.
They think that John wants to go out with Kelly.
Who do they think that wants to go out with Kelly?

# 3.
They think that Jeff bought a house last week.
What do they think that Jeff bought last week?

# 4.
They think that Sarah got upset about the scam.
Who do they think got upset about the scam?

# 5.
They think that Jeff committed the robbery the other day.
Who do they think that committed the robbery the other day?

# 6.
They think that we should choose John as a spokesperson.
Who(m) do they think that we should choose as a spokesperson?

# 7.
They think that John has provided one million dollars for Jane’s project.
Who do they think that has provided one million dollars for Jane’s project?

# 8.
They think that John is going to marry Sarah next month.
Who do they think is going to marry Sarah next month?

# 9.
They think that Meredith likes jazz.
What kind of music do they think that Meredith likes?

# 10.
They think that John saw Robert De Niro in the street.
Who do they think saw Robert De Niro in the street?
# 11.
They think that John broke up with Silva the other day. 
Who do they think that broke up with Silva the other day?

# 12.
They think that Mike has criticized the war in Iraq.
What do they think that Mike has criticized?

# 13.
They believe that John arrested the criminal last Saturday.
Who do they believe arrested the criminal last Saturday?

# 14.
They think that John falls in love with Sarah.
Who do they think that falls in love with Sarah?

# 15.
They think that Sarah raised Jane.
Who(m) do they think that Sarah raised?

# 16.
They think that John should do this important project.
Who do they think that should do this important project?

# 17.
They think that John will approve their financial package.
Who do they think will approve their financial package?

# 18.
They think that John did his math homework last night.
What do they think that John did last night?

# 19.
They think that Sarah should lead their reading group.
Who do they think should lead their reading group?

# 20.
They think that John stopped Anna’s car yesterday.
Who do they think that stopped Anna’s car yesterday?

# 21.
They think that John watches CNN.
Which news channel do they think that John watches?
# 22.
They think that Sarah surprised everyone at the party.  
Who do they think surprised everyone at the party?

# 23.
They think that John should be fired as soon as possible.  
Who do they think that should be fired as soon as possible?

# 24.
They think that Sarah went shopping with Jamie last Friday.  
Who do they think went shopping with Jamie last Friday?

Type B

# 1.
They think that John sells hot dogs.  
What do they think that John sells?

# 2.
They think that John lost his wallet the other day.  
What do they think that John lost the other day?

# 3.
They think that John went to a bar alone yesterday.  
Who do they think that went to a bar alone yesterday?

# 4.
They think that Sarah will provide John with research opportunities.  
Who do they think will provide John with research opportunities?

# 5.
They think that the Kia Tigers will meet the LG Twins in the Korean Series.  
Which team do they think that the Kia Tigers will meet in the Korean Series?

# 6.
They think that John has been investigating the crime since last month.  
Who do they think that has been investigating the crime since last month?

# 7.
They think that John made Italian pizza at Jennifer’s place the other day.  
Who do they think made Italian pizza at Jennifer’s place the other day?
# 8.
They think that John ate pasta at that restaurant.
What do they think that John ate at that restaurant?

# 9.
They think that Sarah has treated John’s brother with respect.
Who do they think has treated John’s brother with respect?

# 10.
They think that Sarah should marry Jeff.
Who do they think that should marry Jeff?

# 11.
They think that John has betrayed our company.
Who do they think has betrayed our company?

# 12.
They think that John passed the exam last Friday?
Who do they think that passed the exam last Friday?

# 13.
They think that Bill studied Japanese last semester.
What/which language do they think that Bill studied last semester?

# 14.
They think that Jerry’s Pizza serves Chicago style pizza.
Which restaurant do they think serves Chicago style pizza?

# 15.
They think that John has been working for Jamie’s brother.
Who do they think that has been working for Jamie’s brother?

# 16.
They think that Michelle watched Shrek 2 the other day.
What/which movie do they think that Michelle watched the other day?

# 17.
They think that John drove Susan to the airport yesterday.
Who do they think that drove Susan to the airport yesterday?

# 18.
They think that Sarah dated Jimmy two years ago.
Who(m) do they think that Sarah dated two years ago?
# 19. They think that John stole Jane’s engagement ring last summer. Who do they think stole Jane’s engagement ring last summer?

# 20. They think that John will allow Jeff to continue his work. Who do they think that will allow Jeff to continue his work?

# 21. They think that John lied to the police about the murder suspect yesterday. Who do they think lied to the police about the murder suspect yesterday?

# 22. They think that John happened to meet Karen this morning. Who do they think that happened to meet Karen this morning?

# 23. They think that John kissed Sarah at the party. Who(m) do they think that John kissed at the party?

# 24. They think that John will watch this year’s Super Bowl with Jane. Who do they think will watch this year’s Super Bowl with Jane?

Type C

# 1. They think that John hit Mike last night. Who(m) do they think that John hit last night?

# 2. They think that Jane met Alan at Mike’s party. Who(m) do they think that Jane met at Mike’s party?

# 3. They think that Jeff bought a laptop yesterday. Who do they think that bought a laptop yesterday?

# 4. They think that Sarah married Mr. Smith last year. Who do they think that married Mr. Smith last year?
# 5.
They think that Mike broke Jane’s car the other day.
What do they think that Mike broke the other day?

# 6.
They think that John should talk to the principal.
Who do they think should talk to the principal?

# 7.
They think that Mary ran a red light last night.
Who do they think ran a red light last night?

# 8.
They think that Mary submitted her report yesterday.
What do they think that Mary submitted yesterday?

# 9.
They think that Mary will become a novelist.
Who do they think will become a novelist?

# 10.
They think that Jennifer had dinner with Bill Clinton.
Who do they think that had dinner with Bill Clinton?

# 11.
They think that Jeff killed a rat the other day.
Who do they think that killed a rat the other day?

# 12.
They think that Obama will meet Kim Jong-il next year.
Who do they think will meet Kim Jong-il next year?

# 13.
They think that Gary stole Jane’s diamond last night.
What do they think that Gary stole last night?

# 14.
They think that John borrowed $200 from Jimmy.
Who do they think borrowed $200 from Jimmy?

# 15.
They think that Sarah got drunk the other day.
Who do they think got drunk the other day?

# 16.
They think that Alan will have a birthday party soon.
Who do they think that will have a birthday party soon?
# 17.
They think that Al Pacino is the most talented actor.
Who do they think is the most talented actor?

# 18.
They think that Jennifer saw the movie “Avatar” last night.
What movie do they think that Jennifer saw last night?

# 19.
They think that John hired Meredith last year.
Who do they think that hired Meredith last year?

# 20.
They think that Alex is going to visit Jim next week.
Who(m) do they think that Alex is going to visit next week?

# 21.
They think that John ran a marathon last spring.
Who do they think that ran a marathon last spring?

# 22.
They think that Mike hated clam chowder at the party.
What do they think that Mike hated at the party?

# 23.
They think that Jeff is responsible for the robbery.
Who do they think that is responsible for the robbery?

# 24.
They think that Maria went swimming with Jim yesterday.
Who do they think went swimming with Jim yesterday?
Appendix D

Grammaticality Judgment Test Items (Past unreal conditional)

Type A

# 1.
He didn’t pay his phone bill. He could not use his phone.
If he had paid his phone bill, he could use his phone.

# 2.
He did not take care of his dogs. They got sick.
If he had taken care of his dogs, they wouldn’t have gotten sick.

# 3.
Sonia did not save a lot of money last year. She could not afford a new car.
If Sonia had saved a lot of money last year, she could have afforded a new car.

# 4.
He was not nice to her. She did not want to talk to him.
If he had been nice to her, she would/might have wanted to talk to him.

# 5.
I cheated on my wife last year. I felt guilty.
If I did not cheat on my wife last year, I wouldn’t have felt guilty.

# 6.
Jane had a car accident. She could not play soccer.
If Jane didn’t have a car accident, she could have played soccer.

# 7.
Jane did not work hard last year. She was not a good candidate for the job.
If Jane worked hard last year, she would/could be a good candidate for the job.

# 8.
Jeff fired Mary. She had to find a new job.
If Jeff did not fire Mary, she would not have to find a new job.

# 9.
The teacher postponed the test. I had enough time to prepare for it.
If the teacher had not postponed the test, I wouldn’t have had enough time to prepare for it.

# 10.
His company prospered last year. He was happy.
If his company had not prospered last year, he would not be happy.
# 11.
I broke my arm. I could not play golf.
If I had not broken my arm, I could have played golf.

# 12.
I did not close the window. I had a cold.
If I closed the window, I would not have a cold.

# 13.
He lived in Spain for 25 years. He became a master of Spanish.
If he had not lived in Spain for 25 years, he would/might not have become a master of Spanish.

# 14.
He didn’t come to the class. He didn’t know what to do for his final project.
If he came to the class, he would know what to do for his final project.

# 15.
I had a huge breakfast. I skipped lunch.
If I had not had a huge breakfast, I would/might not have skipped lunch.

# 16.
John made a lot of money last year. He could buy an airplane.
If John had not made a lot of money last year, he could not have bought an airplane (or would not have been able to buy an airplane).

Type B

# 1.
The chairman did not attend the meeting. They had another meeting the following day.
If the chairman had attended the meeting, they wouldn’t have had another meeting the following day.

# 2.
I didn’t sleep well last night. I fell asleep this morning.
If I had slept well last night, I would/might not have fallen asleep this morning.

# 3.
The director made that movie. My friends were able to recognize him.
If the director had not made that movie, my friends would not be able to recognize him.

# 4.
John got fired. He did not receive his paycheck last week.
If John didn’t get fired, he would have received his paycheck last week.
# 5.
John called the police immediately. The police could arrest the criminal at the airport.
If John had not called the police immediately, the police could not arrest the criminal at the airport.

# 6.
She didn’t answer my call. I decided to break up with her.
If she had answered my call, I would not have decided to break up with her.

# 7.
The plane did not arrive on time. I had to run to the meeting.
If the plane arrived on time, I wouldn’t have to run to the meeting.

# 8.
He didn’t spend his money wisely. He became penniless.
If he spent his money wisely, he would not become penniless.

# 9.
Dora did not finish faxing all the letters. She could not take a day off.
If Dora had finished faxing all the letters, she could have taken a day off.

# 10.
Sarah spent $10,000 on clothes. She could not pay her tuition last year.
If Sarah did not spend $10,000 on clothes, she could pay her tuition last year.

# 11.
I lost my passport. I asked for a new one.
If I had not lost my passport, I wouldn’t have asked for a new one.

# 12.
I had a lot of beer. I had a hangover.
If I had not had a lot of beer, I wouldn’t have had a hangover.

# 13.
I didn’t pay my credit card bills. I could not use my credit card.
If I had paid my credit card bills, I could have used my credit card.

# 14.
I stopped for a cup of coffee. I was late for the class.
If I didn’t stop for a cup of coffee, I would not be late for the class.

# 15.
I ignored traffic signals this morning. I got a ticket.
If I did not ignore traffic signals this morning, I wouldn’t have gotten a ticket.
# 16.
He went to Chicago yesterday. He couldn’t help me with my homework.
If he hadn’t gone to Chicago yesterday, he could have helped me with my homework.

Type C

# 1.
Mary started her own company. She was able to support her parents.
If Mary did not start her own company, she would not have been able to support her parents.

# 2.
He was not in the U.S. last year. I couldn’t contact him.
If he had been in the U.S. last year, I could have contacted him.

# 3.
He broke his ankle last night. He did not work out this morning.
If he did not break his ankle last night, he would work out this morning.

# 4.
Mary did not give me her phone number. I couldn’t invite her to the banquet the other day.
If Mary had given me her phone number, I could/would invite her to the banquet the other day.

# 5.
I had too much coffee. I could not go to bed early last night.
If I did not have too much coffee, I could/might go to bed early last night.

# 6.
I didn’t have a job last year. I didn’t have to get up early in the morning.
If I had had a job last year, I would have to get up early in the morning.

# 7.
I spent too much time baby-sitting. I did not have enough time to finish my homework.
If I had not spent too much time baby-sitting, I would have had enough time to finish my homework.

# 8.
Sarah was not careful. She had a car accident.
If Sarah had been (more) careful, she would/might not have had a car accident.

# 9.
Angela didn’t have a phone. Jeff was not able to give her a wake-up call.
If Angela had a phone, Jeff would have been able to give her a wake-up call.
# 10.
He didn’t pass the bar exam. His girlfriend left him.
If he had passed the bar exam, his girlfriend wouldn’t have left him.

# 11.
They were attracted to each other at the party. They began dating.
If they had not been attracted to each other at the party, they wouldn’t have begun dating.

# 12.
Alex finished his work on time. The company did not issue him a warning.
If Alex didn’t finish his work on time, the company would issue him a warning.

# 13.
I did not talk back to my boss. I was able to keep my job.
If I had talked back to my boss, I would not have been able to keep my job.

# 14.
He didn’t have a car. He used public transportation.
If he had a car, he would (might) not use public transportation.

# 15.
She lost her key to her office. She was not able to access her computer.
If she had not lost her key to her office, she would have been able to access her computer.

# 16.
I didn’t have enough money. I could not buy a new car for my son last year.
If I had had enough money, I could have bought a new car for my son last year.
Appendix E
Grammaticality Judgment Test

School: ___________________ Name: ___________________ Type: _______

Directions: In this task, you need to judge the grammaticality of the interrogative presented on each slide for Part 1 and that of the conditional sentence presented for Part 2 and then put a check mark (√) in the appropriate box within 10 seconds for each item in Part 1 and 15 seconds for each item in Part 2 (grammatical, ungrammatical, or I don’t know). Please do NOT take a guess. If you don’t know, put a check mark in the “I don’t know” box.

<table>
<thead>
<tr>
<th>Part 1 (Questions)</th>
<th>Part 2 (Conditional)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grammatical</td>
<td>Ungrammatical</td>
</tr>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td></td>
</tr>
</tbody>
</table>
Please correct the sentences that you judged ungrammatical (Part 1)

School: _______________  Name: ___________________________  Type: _______

Sentence # (    )

______________________________

Sentence # (    )

______________________________

Sentence # (    )

______________________________

Sentence # (    )

______________________________

Sentence # (    )

______________________________

Sentence # (    )

______________________________

Sentence # (    )

______________________________

Sentence # (    )

______________________________
Please correct the sentences that you judged ungrammatical (Part 2)

School: ________________  Name: ___________________________  Type: ______

Sentence # (   )

____________________________________________________________

Sentence # (   )

____________________________________________________________

Sentence # (   )

____________________________________________________________

Sentence # (   )

____________________________________________________________

Sentence # (   )

____________________________________________________________

Sentence # (   )

____________________________________________________________

Sentence # (   )

____________________________________________________________
Appendix F

GJT Performance on *That*-Trace Filter Regressed on Cognitive Variables

Table F1

*GJT Performance of Experimental Groups Regressed on Cognitive Measures*

<table>
<thead>
<tr>
<th>Variable</th>
<th>$B$</th>
<th>$SE B$</th>
<th>$\beta$</th>
<th>$t$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Posttest</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WMC</td>
<td>-9.157</td>
<td>10.138</td>
<td>-0.158</td>
<td>-0.903</td>
<td>0.373</td>
</tr>
<tr>
<td>RT cost</td>
<td>0.000</td>
<td>0.004</td>
<td>-0.016</td>
<td>-0.089</td>
<td>0.930</td>
</tr>
<tr>
<td>Error cost</td>
<td>-0.048</td>
<td>0.186</td>
<td>-0.045</td>
<td>-0.256</td>
<td>0.799</td>
</tr>
<tr>
<td><strong>Delayed Posttest</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WMC</td>
<td>-14.557</td>
<td>9.830</td>
<td>-0.253</td>
<td>-1.481</td>
<td>0.148</td>
</tr>
<tr>
<td>RT cost</td>
<td>0.000</td>
<td>0.004</td>
<td>-0.018</td>
<td>-0.103</td>
<td>0.918</td>
</tr>
<tr>
<td>Error cost</td>
<td>-0.111</td>
<td>0.190</td>
<td>-0.102</td>
<td>-0.582</td>
<td>0.564</td>
</tr>
</tbody>
</table>

*Note.* The values were based on the data collected from those who participated in all three test sessions. $B$ means an unstandardized regression coefficient (regression slope). $SE B$ means a standard error of $B$. $\beta$ means a standardized coefficient, which is the same as a correlation coefficient in a simple linear regression analysis. A $t$-value tests how a given variable contributes to a corresponding regression model.

Table F2

*GJT Performance of Experimental Groups Regressed on Cognitive Measures (Posttest Only)*

<table>
<thead>
<tr>
<th>Variable</th>
<th>$B$</th>
<th>$SE B$</th>
<th>$\beta$</th>
<th>$T$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>WMC</td>
<td>-4.951</td>
<td>8.537</td>
<td>-0.096</td>
<td>-0.580</td>
<td>0.566</td>
</tr>
<tr>
<td>RT cost</td>
<td>0.001</td>
<td>0.004</td>
<td>0.028</td>
<td>0.168</td>
<td>0.868</td>
</tr>
<tr>
<td>Error cost</td>
<td>-.027</td>
<td>.180</td>
<td>-.025</td>
<td>-.150</td>
<td>.882</td>
</tr>
</tbody>
</table>

*Note.* The values were based on the data collected from those who attended the pretest and the posttest only.
Table F3

*Gain Scores Regressed on Cognitive Measures (Experimental Groups Combined)*

<table>
<thead>
<tr>
<th>Variable</th>
<th>B</th>
<th>SE B</th>
<th>β</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pre-to-Post Gain Scores</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WMC</td>
<td>-.140</td>
<td>.792</td>
<td>-.029</td>
<td>-.177</td>
<td>.861</td>
</tr>
<tr>
<td>RT cost</td>
<td>6.343E-5</td>
<td>.000</td>
<td>.030</td>
<td>.184</td>
<td>.855</td>
</tr>
<tr>
<td>Error cost</td>
<td>.008</td>
<td>.018</td>
<td>.076</td>
<td>.462</td>
<td>.647</td>
</tr>
<tr>
<td><strong>Pre-to-Delayed Gain Scores</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WMC</td>
<td>-1.014</td>
<td>.858</td>
<td>-.205</td>
<td>-1.182</td>
<td>.246</td>
</tr>
<tr>
<td>RT cost</td>
<td>5.606E-5</td>
<td>.000</td>
<td>.027</td>
<td>.155</td>
<td>.878</td>
</tr>
<tr>
<td>Error cost</td>
<td>.000</td>
<td>.017</td>
<td>.005</td>
<td>.028</td>
<td>.978</td>
</tr>
</tbody>
</table>

Table F4

*Pre-to-Post Gain Scores in GJT on That-Trace Regressed on Cognitive Measures*

<table>
<thead>
<tr>
<th>Variable</th>
<th>B</th>
<th>SE B</th>
<th>β</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Recast</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WMC</td>
<td>.103</td>
<td>1.039</td>
<td>.023</td>
<td>.099</td>
<td>.922</td>
</tr>
<tr>
<td>RT cost</td>
<td>.000</td>
<td>.000</td>
<td>-.211</td>
<td>-.942</td>
<td>.358</td>
</tr>
<tr>
<td>Error cost</td>
<td>-.034</td>
<td>.025</td>
<td>-.296</td>
<td>-1.352</td>
<td>.192</td>
</tr>
<tr>
<td><strong>Metalinguistic</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WMC</td>
<td>-1.185</td>
<td>1.203</td>
<td>-.247</td>
<td>-.985</td>
<td>.340</td>
</tr>
<tr>
<td>RT cost</td>
<td>.001</td>
<td>.001</td>
<td>.272</td>
<td>1.132</td>
<td>.274</td>
</tr>
<tr>
<td>Error cost</td>
<td>.045</td>
<td>.022</td>
<td>.459</td>
<td>2.067</td>
<td>.055</td>
</tr>
</tbody>
</table>

193
Table F5

*Pre-to-Delayed Gain Scores in GJT Regressed on Cognitive Measures*

<table>
<thead>
<tr>
<th>Variable</th>
<th>$B$</th>
<th>$SE$ $B$</th>
<th>$\beta$</th>
<th>$t$</th>
<th>$P$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Recast</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WMC</td>
<td>-.163</td>
<td>1.247</td>
<td>-.032</td>
<td>-.131</td>
<td>.897</td>
</tr>
<tr>
<td>RT cost</td>
<td>.000</td>
<td>.000</td>
<td>-.142</td>
<td>-.591</td>
<td>.562</td>
</tr>
<tr>
<td>Error cost</td>
<td>-.034</td>
<td>.025</td>
<td>-.305</td>
<td>-1.320</td>
<td>.204</td>
</tr>
<tr>
<td><strong>Metalinguistic</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WMC</td>
<td>-2.085</td>
<td>1.180</td>
<td>-.440</td>
<td>-1.768</td>
<td>.101</td>
</tr>
<tr>
<td>RT cost</td>
<td>.000</td>
<td>.001</td>
<td>.209</td>
<td>.769</td>
<td>.456</td>
</tr>
<tr>
<td>Error cost</td>
<td>.030</td>
<td>.023</td>
<td>.339</td>
<td>1.300</td>
<td>.216</td>
</tr>
</tbody>
</table>
Appendix G

Oral Performance on *That*-Trace filter Regressed on Cognitive Variables

Table G1

*Posttest Performance Regressed on Cognitive Measures (Experimental Groups Combined)*

<table>
<thead>
<tr>
<th>Variable</th>
<th>B</th>
<th>SE B</th>
<th>β</th>
<th>T</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>WMC</td>
<td>3.592</td>
<td>8.714</td>
<td>.070</td>
<td>.412</td>
<td>.683</td>
</tr>
<tr>
<td>RT cost</td>
<td>.003</td>
<td>.004</td>
<td>.106</td>
<td>.622</td>
<td>.538</td>
</tr>
<tr>
<td>Error cost</td>
<td>-.164</td>
<td>.207</td>
<td>-.135</td>
<td>-.792</td>
<td>.434</td>
</tr>
</tbody>
</table>

Table G2

*Gain Scores Regressed on Cognitive Measures (Experimental Groups Combined)*

<table>
<thead>
<tr>
<th>Variable</th>
<th>B</th>
<th>SE B</th>
<th>β</th>
<th>T</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>WMC</td>
<td>.321</td>
<td>.777</td>
<td>.070</td>
<td>.413</td>
<td>.682</td>
</tr>
<tr>
<td>RT cost</td>
<td>.000</td>
<td>.000</td>
<td>.118</td>
<td>.692</td>
<td>.493</td>
</tr>
<tr>
<td>Error cost</td>
<td>-.014</td>
<td>.019</td>
<td>-.125</td>
<td>-.735</td>
<td>.467</td>
</tr>
</tbody>
</table>

Pre-to-Post Gain Scores

Pre-to-Delayed Gain Scores

| WMC        | -.174 | .975 | -.032 | -.179 | .859 |
| RT cost    | .001 | .000 | .290 | 1.634 | .113 |
| Error cost | -.013 | .021 | -.118 | -.638 | .529 |
Table G3

_Pre-to-Post Gain Scores Regressed on Cognitive Measures by Group_

<table>
<thead>
<tr>
<th>Variable</th>
<th>$B$</th>
<th>SE $B$</th>
<th>$\beta$</th>
<th>$T$</th>
<th>$P$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recast</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WMC</td>
<td>.750</td>
<td>1.323</td>
<td>.132</td>
<td>.567</td>
<td>.578</td>
</tr>
<tr>
<td>RT cost</td>
<td>.000</td>
<td>.001</td>
<td>.123</td>
<td>.510</td>
<td>.617</td>
</tr>
<tr>
<td>Error cost</td>
<td>-.022</td>
<td>.035</td>
<td>-.149</td>
<td>-.620</td>
<td>.544</td>
</tr>
<tr>
<td>Metalinguistic</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WMC</td>
<td>-.932</td>
<td>.634</td>
<td>-.355</td>
<td>-1.469</td>
<td>.162</td>
</tr>
<tr>
<td>RT cost</td>
<td>.000</td>
<td>.000</td>
<td>.130</td>
<td>.506</td>
<td>.620</td>
</tr>
<tr>
<td>Error cost</td>
<td>.004</td>
<td>.015</td>
<td>.060</td>
<td>.232</td>
<td>.820</td>
</tr>
</tbody>
</table>

Table G4

_Pre-to-Delayed Gain Scores Regressed on Cognitive Measures by Group_

<table>
<thead>
<tr>
<th>Variable</th>
<th>$B$</th>
<th>SE $B$</th>
<th>$\beta$</th>
<th>$T$</th>
<th>$P$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recast</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WMC</td>
<td>.623</td>
<td>1.607</td>
<td>.096</td>
<td>.387</td>
<td>.704</td>
</tr>
<tr>
<td>RT cost</td>
<td>.000</td>
<td>.001</td>
<td>.147</td>
<td>.574</td>
<td>.574</td>
</tr>
<tr>
<td>Error cost</td>
<td>-.045</td>
<td>.036</td>
<td>-.311</td>
<td>-1.267</td>
<td>.224</td>
</tr>
<tr>
<td>Metalinguistic</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WMC</td>
<td>-1.251</td>
<td>1.039</td>
<td>-.317</td>
<td>-1.204</td>
<td>.250</td>
</tr>
<tr>
<td>RT cost</td>
<td>.001</td>
<td>.000</td>
<td>.518</td>
<td>2.095</td>
<td>.058</td>
</tr>
<tr>
<td>Error cost</td>
<td>.018</td>
<td>.023</td>
<td>.221</td>
<td>.786</td>
<td>.447</td>
</tr>
</tbody>
</table>
Appendix H

GJT Performance on Past Unreal Conditional Regressed on Cognitive Variables

Table H1

*Performance of Experimental Groups Regressed on Cognitive Measures*

<table>
<thead>
<tr>
<th>Variable</th>
<th>B</th>
<th>SE B</th>
<th>β</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Posttest</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WMC</td>
<td>3.696</td>
<td>5.081</td>
<td>.139</td>
<td>.727</td>
<td>.473</td>
</tr>
<tr>
<td>RT cost</td>
<td>.002</td>
<td>.003</td>
<td>.194</td>
<td>.971</td>
<td>.341</td>
</tr>
<tr>
<td>Error cost</td>
<td>.110</td>
<td>.110</td>
<td>.199</td>
<td>.996</td>
<td>.329</td>
</tr>
<tr>
<td><strong>Delayed Posttest</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WMC</td>
<td>1.183</td>
<td>6.908</td>
<td>.033</td>
<td>.171</td>
<td>.865</td>
</tr>
<tr>
<td>RT cost</td>
<td>.000</td>
<td>.003</td>
<td>.009</td>
<td>.042</td>
<td>.967</td>
</tr>
<tr>
<td>Error cost</td>
<td>.176</td>
<td>.139</td>
<td>.250</td>
<td>1.264</td>
<td>.218</td>
</tr>
</tbody>
</table>

Table H2

*Posttest Performance Regressed on Cognitive Measures (Experimental Groups Combined)*

<table>
<thead>
<tr>
<th>Variable</th>
<th>B</th>
<th>SE B</th>
<th>β</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>WMC</td>
<td>1.806</td>
<td>4.377</td>
<td>.072</td>
<td>.413</td>
<td>.683</td>
</tr>
<tr>
<td>RT cost</td>
<td>.000</td>
<td>.002</td>
<td>.033</td>
<td>.183</td>
<td>.856</td>
</tr>
<tr>
<td>Error cost</td>
<td>.079</td>
<td>.098</td>
<td>.144</td>
<td>.809</td>
<td>.425</td>
</tr>
</tbody>
</table>
Table H3

*Gain Scores Regressed on Cognitive Measures (Experimental Groups Combined)*

<table>
<thead>
<tr>
<th>Variable</th>
<th>$B$</th>
<th>$SE,B$</th>
<th>$\beta$</th>
<th>$t$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pre-to-Post Gain Scores</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WMC</td>
<td>.464</td>
<td>.544</td>
<td>.147</td>
<td>.852</td>
<td>.401</td>
</tr>
<tr>
<td>RT cost</td>
<td>.000</td>
<td>.000</td>
<td>.139</td>
<td>.784</td>
<td>.439</td>
</tr>
<tr>
<td>Error cost</td>
<td>.015</td>
<td>.013</td>
<td>.203</td>
<td>1.152</td>
<td>.258</td>
</tr>
<tr>
<td></td>
<td>Pre-to-Delayed Gain Scores</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WMC</td>
<td>-.031</td>
<td>1.161</td>
<td>-.005</td>
<td>-.026</td>
<td>.979</td>
</tr>
<tr>
<td>RT cost</td>
<td>1.551E-5</td>
<td>.001</td>
<td>.006</td>
<td>.027</td>
<td>.979</td>
</tr>
<tr>
<td>Error cost</td>
<td>.039</td>
<td>.023</td>
<td>.324</td>
<td>1.679</td>
<td>.106</td>
</tr>
</tbody>
</table>

Table H4

*Pre-to-Post Gain Scores Regressed on Cognitive Measures by Group*

<table>
<thead>
<tr>
<th>Variable</th>
<th>$B$</th>
<th>$SE,B$</th>
<th>$\beta$</th>
<th>$t$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Recast</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WMC</td>
<td>.095</td>
<td>.869</td>
<td>.027</td>
<td>.109</td>
<td>.914</td>
</tr>
<tr>
<td>RT cost</td>
<td>.000</td>
<td>.000</td>
<td>.077</td>
<td>.307</td>
<td>.763</td>
</tr>
<tr>
<td>Error cost</td>
<td>.016</td>
<td>.017</td>
<td>.229</td>
<td>.941</td>
<td>.361</td>
</tr>
<tr>
<td></td>
<td>Metalinguistic</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WMC</td>
<td>.876</td>
<td>.810</td>
<td>.278</td>
<td>1.082</td>
<td>.297</td>
</tr>
<tr>
<td>RT cost</td>
<td>.000</td>
<td>.000</td>
<td>.234</td>
<td>.869</td>
<td>.400</td>
</tr>
<tr>
<td>Error cost</td>
<td>.014</td>
<td>.023</td>
<td>.168</td>
<td>.615</td>
<td>.549</td>
</tr>
</tbody>
</table>
Table H5

*Pre-to-Delayed Gain Scores Regressed on Cognitive Measures by Group*

<table>
<thead>
<tr>
<th>Variable</th>
<th>B</th>
<th>SE B</th>
<th>β</th>
<th>t</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Recast</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WMC</td>
<td>-.030</td>
<td>1.524</td>
<td>-.005</td>
<td>-.019</td>
<td>.985</td>
</tr>
<tr>
<td>RT cost</td>
<td>.000</td>
<td>.001</td>
<td>.186</td>
<td>.627</td>
<td>.543</td>
</tr>
<tr>
<td>Error cost</td>
<td>.044</td>
<td>.025</td>
<td>.474</td>
<td>1.786</td>
<td>.102</td>
</tr>
<tr>
<td><strong>Metalinguistic</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WMC</td>
<td>-.618</td>
<td>2.101</td>
<td>-.085</td>
<td>-.294</td>
<td>.774</td>
</tr>
<tr>
<td>RT cost</td>
<td>.000</td>
<td>.001</td>
<td>-.144</td>
<td>-.482</td>
<td>.639</td>
</tr>
<tr>
<td>Error cost</td>
<td>.040</td>
<td>.048</td>
<td>.244</td>
<td>.835</td>
<td>.421</td>
</tr>
</tbody>
</table>
Appendix I

Oral Performance on Past Unreal Conditional Regressed on Cognitive Variables

Table I1

Performance of Experimental Groups Regressed on Cognitive Measures

<table>
<thead>
<tr>
<th>Variable</th>
<th>B</th>
<th>SE B</th>
<th>β</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Posttest</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WMC</td>
<td>11.535</td>
<td>4.997</td>
<td>.400</td>
<td>2.308*</td>
<td>.029</td>
</tr>
<tr>
<td>RT cost</td>
<td>.000</td>
<td>.003</td>
<td>.020</td>
<td>.099</td>
<td>.922</td>
</tr>
<tr>
<td>Error cost</td>
<td>.017</td>
<td>.116</td>
<td>.029</td>
<td>.147</td>
<td>.885</td>
</tr>
<tr>
<td>Delayed Posttest</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WMC</td>
<td>6.921</td>
<td>9.680</td>
<td>.134</td>
<td>.715</td>
<td>.481</td>
</tr>
<tr>
<td>RT cost</td>
<td>.005</td>
<td>.004</td>
<td>.234</td>
<td>1.203</td>
<td>.240</td>
</tr>
<tr>
<td>Error cost</td>
<td>.243</td>
<td>.193</td>
<td>.244</td>
<td>1.260</td>
<td>.219</td>
</tr>
</tbody>
</table>

*p < .05.

Table I2

Posttest Performance Regressed on Cognitive Measures (Experimental Groups Combined)

<table>
<thead>
<tr>
<th>Variable</th>
<th>B</th>
<th>SE B</th>
<th>β</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>WMC</td>
<td>8.819</td>
<td>4.542</td>
<td>.316</td>
<td>1.942</td>
<td>.060</td>
</tr>
<tr>
<td>RT cost</td>
<td>7.045E-6</td>
<td>.002</td>
<td>.001</td>
<td>.003</td>
<td>.998</td>
</tr>
<tr>
<td>Error cost</td>
<td>.010</td>
<td>.103</td>
<td>.016</td>
<td>.093</td>
<td>.926</td>
</tr>
</tbody>
</table>
### Table I3

*Gain Scores Regressed on Cognitive Measures (Experimental Groups Combined)*

<table>
<thead>
<tr>
<th>Variable</th>
<th>$B$</th>
<th>$SE B$</th>
<th>$\beta$</th>
<th>$t$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pre-to-Post Development</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WMC</td>
<td>.834</td>
<td>.396</td>
<td>.340</td>
<td>2.108*</td>
<td>.042</td>
</tr>
<tr>
<td>RT cost</td>
<td>-5.340E-6</td>
<td>.000</td>
<td>-.005</td>
<td>-.026</td>
<td>.979</td>
</tr>
<tr>
<td>Error cost</td>
<td>.000</td>
<td>.009</td>
<td>-.004</td>
<td>-.023</td>
<td>.982</td>
</tr>
<tr>
<td><strong>Pre-to-Delayed Development</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WMC</td>
<td>.832</td>
<td>.822</td>
<td>.188</td>
<td>1.012</td>
<td>.320</td>
</tr>
<tr>
<td>RT cost</td>
<td>.000</td>
<td>.000</td>
<td>.185</td>
<td>.939</td>
<td>.357</td>
</tr>
<tr>
<td>Error cost</td>
<td>.019</td>
<td>.017</td>
<td>.217</td>
<td>1.112</td>
<td>.277</td>
</tr>
</tbody>
</table>

* $p < .05.$

### Table I4

*Pre-to-Post Gain Scores Regressed on Cognitive Measures by Group*

<table>
<thead>
<tr>
<th>Variable</th>
<th>$B$</th>
<th>$SE B$</th>
<th>$\beta$</th>
<th>$t$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Recast</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WMC</td>
<td>.078</td>
<td>.563</td>
<td>.034</td>
<td>.139</td>
<td>.891</td>
</tr>
<tr>
<td>RT cost</td>
<td>-.001</td>
<td>.000</td>
<td>-.506</td>
<td>-2.275*</td>
<td>.038</td>
</tr>
<tr>
<td>Error cost</td>
<td>-.001</td>
<td>.011</td>
<td>-.035</td>
<td>-.135</td>
<td>.894</td>
</tr>
<tr>
<td><strong>Metalinguistic</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WMC</td>
<td>1.733</td>
<td>.578</td>
<td>.612</td>
<td>2.996*</td>
<td>.009</td>
</tr>
<tr>
<td>RT cost</td>
<td>.000</td>
<td>.000</td>
<td>.363</td>
<td>1.511</td>
<td>.152</td>
</tr>
<tr>
<td>Error cost</td>
<td>.008</td>
<td>.017</td>
<td>.120</td>
<td>.469</td>
<td>.646</td>
</tr>
</tbody>
</table>

* $p < .05.$
Table 15

*Pre-to-Delayed Gain Scores Regressed on Cognitive Measures by Group*

<table>
<thead>
<tr>
<th>Variable</th>
<th>$B$</th>
<th>$SE B$</th>
<th>$\beta$</th>
<th>$t$</th>
<th>$P$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Recast</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WMC</td>
<td>-.258</td>
<td>1.163</td>
<td>-.062</td>
<td>-.222</td>
<td>.828</td>
</tr>
<tr>
<td>RT cost</td>
<td>.001</td>
<td>.001</td>
<td>.410</td>
<td>1.423</td>
<td>.185</td>
</tr>
<tr>
<td>Error cost</td>
<td>.018</td>
<td>.022</td>
<td>.260</td>
<td>.850</td>
<td>.415</td>
</tr>
<tr>
<td><strong>Metalinguistic</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WMC</td>
<td>2.761</td>
<td>1.188</td>
<td>.542</td>
<td>2.325*</td>
<td>.037</td>
</tr>
<tr>
<td>RT cost</td>
<td>8.093E-5</td>
<td>.001</td>
<td>.040</td>
<td>.145</td>
<td>.887</td>
</tr>
<tr>
<td>Error cost</td>
<td>.022</td>
<td>.030</td>
<td>.198</td>
<td>.730</td>
<td>.479</td>
</tr>
</tbody>
</table>

*p < .05.*
References


Oxford University Press.


doi:10.1016/j.cognition.2009.03.001


Studies in Second Language Acquisition, 18, 91-126.


doi:10.1111/j.1467-9922.2010.00573.x


Learning, 59, 453-498.


London: Lawrence Erlbaum Associates.


learning (pp. 211-266). Amsterdam: John Benjamins.


Sagarra, N. (2007). From CALL to face-to-face interaction: The effect of computer-delivered recasts and working memory on L2 development. In A. Mackey (Ed.), *Conversational...


doi:10.1017/S0272263109990507


doi:10.1017/S0272263109990532


Trofimovich, P., Ammar, A., & Gatbonton, E. (2007). How effective are recasts? The role of attention, memory, and analytic ability. In A. Mackey (Ed.), Conversational interaction in second language acquisition: A collection of empirical studies (pp. 171-


