http://researchspace.auckland.ac.nz

University of Auckland Research Repository, ResearchSpace

Copyright Statement

The digital copy of this thesis is protected by the Copyright Act 1994 (New Zealand).

This thesis may be consulted by you, provided you comply with the provisions of the Act and the following conditions of use:

- Any use you make of these documents or images must be for research or private study purposes only, and you may not make them available to any other person.
- Authors control the copyright of their thesis. You will recognise the author's right to be identified as the author of this thesis, and due acknowledgement will be made to the author where appropriate.
- You will obtain the author's permission before publishing any material from their thesis.

General copyright and disclaimer

In addition to the above conditions, authors give their consent for the digital copy of their work to be used subject to the conditions specified on the Library Thesis Consent Form and Deposit Licence.
The Effects of Explicit and Implicit Recasts on the Acquisition of Two Grammatical Structures and the Mediating Role of Working Memory

by

Yongbin Zhao

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy in Applied Linguistics

The University of Auckland

2015
Abstract

The efficacy of recasts – a strategy for providing corrective feedback – is a matter of controversy in the field of second language acquisition (SLA). According to one position recasts are effective because they provide learners with both negative and positive evidence. However, according to another position they are ineffective because they frequently fail to push learners to correct their own errors. SLA researchers have also disputed the value of recasts on the grounds that they constitute an implicit type of corrective feedback that learners fail to attend to. However, as other researchers have pointed out, recasts are not invariably implicit; in many cases their corrective force is transparent.

This thesis compared the effects of corrective recasts (i.e. repetition plus recasting with stress) that are relatively explicit and implicit recasts (i.e. simple recasting) on the acquisition of two English grammatical structures (3rd person -s and embedded questions). The effects of these two types of recasts were examined in terms of both learners’ acquisition of implicit knowledge (measured by means of elicited imitation and oral production) and their acquisition of explicit knowledge (measured by means of written production and an untimed grammaticality judgment test). It also investigated the role of uptake-with-repair and working memory in recasts-driven L2 learning.

This was a classroom-based quasi-experimental study, with four groups of EFL learners (n = 109) in a Chinese university (i.e. a corrective recast group, an implicit recast group, a task control group, and a test control group). All groups completed a pretest, an immediate posttest, and a delayed posttest. The treatment for the experimental groups consisted of corrective feedback (i.e. corrective recasts and implicit recasts) provided online while the learners performed focused tasks designed to provide contexts for the use of the target structure. The task-based lessons were audio-recorded and transcribed, enabling all occasions where uptake occurred following corrective feedback to be noted. All groups also completed working memory tests.

Language knowledge scores were compared by means of parametric/non-parametric ANOVAs or ANCOVAs to examine the effect of feedback. Correlational analyses were conducted to investigate the relationships between repair rate, knowledge gains and working memory capacity.
Simply performing the tasks without any feedback benefited L2 acquisition, as shown by the fact that the task control group outperformed the test control group in gains in both implicit and explicit knowledge with a large effect size. Corrective recasts facilitated the development of implicit and explicit knowledge but implicit recasts only led to significant gains in explicit knowledge. Corrective recasts were more effective than implicit recasts on the implicit knowledge of embedded questions and the explicit knowledge of 3rd person -s. These results suggest that explicitness in recasts makes a difference to their efficacy, and the superior effect of corrective recasts over implicit recasts on the development of implicit or explicit knowledge depends on the target structure. This relative effect on explicit knowledge was evident for those learners who had low prior knowledge of the target structure, pointing to the mediation of learner proficiency. The repair rate was also higher after corrective recasts than implicit recasts. However, whether repair did or did not occur was found to be generally irrelevant to L2 learning through recasts. Neither phonological short-term memory (tapped by a non-word span test) nor complex working memory (tapped by a listening span test) was, by and large, a decisive factor in recasts-driven L2 learning in this classroom-based study.

The thesis concludes with a discussion of the theoretical and pedagogical implications of the study. I propose that in the context of Chinese university classroom L2 acquisition can be best promoted by means of focused tasks and corrective feedback consisting of explicit recasts, for low proficiency learners in particular.
Acknowledgements

This dissertation could not have been completed without the encouragement and support of a number of people.

First of all, I express my deepest gratitude to Distinguished Professor Rod Ellis, my main supervisor, for his expert guidance, enlightening comments and, continuous support across the entire tenure of my doctoral research. I appreciate very much his prompt and insightful feedback on my numerous drafts of each chapter, and his great patience with my progression.

I am grateful to Dr. Shaofeng Li, my co-supervisor, for his comments on my research proposal, suggestions on statistical analyses and various encouragements.

I am indebted to the 109 participants, who each contributed about 20 hours to this research. The majority of the participants successfully completed 24 language tests (4 types by 3 times by 2 target structures) and 3 working memory tests in addition to 6 instructional tasks.

My thanks also go to Tina Lin and Bianca Chen for coding and scoring some of the data, and to Russell Greenwood for proofreading this dissertation.

I acknowledge that my doctoral research was conducted with the assistance of a University of Auckland Doctoral scholarship.

Last but not least, I would like to thank my family members for their loving support. My parents and late uncle-in-law made my tertiary education possible. My wife supported me whole-heartedly, doing all the housework, and my daughter cheered me up during the arduous writing stage.
Table of Contents

ABSTRACT ......................................................................................................................................... III
ACKNOWLEDGEMENTS .................................................................................................................... V
TABLE OF CONTENTS ....................................................................................................................... VI
LIST OF TABLES ............................................................................................................................... XII
LIST OF FIGURES ............................................................................................................................. XV
ABBREVIATIONS ............................................................................................................................ XVI

CHAPTER 1 INTRODUCTION ............................................................................................................. 1
  1.1 Aim of the study ............................................................................................................................ 1
  1.2 Theoretical motivation .................................................................................................................. 2
  1.3 Practical concerns ......................................................................................................................... 4
  1.4 Structure of the thesis .................................................................................................................. 5

CHAPTER 2 THEORETICAL ISSUES ..................................................................................................... 7
  2.1 Corrective feedback ....................................................................................................................... 7
    2.1.1 Types of corrective feedback ................................................................................................. 7
    2.1.2 Types of recasts ....................................................................................................................... 8
    2.1.3 Positive and negative evidence ............................................................................................ 10
  2.2 Theoretical basis for corrective feedback ................................................................................... 11
    2.2.1 Opponents of corrective feedback ....................................................................................... 12
    2.2.2 Proponents of corrective feedback ....................................................................................... 12
    2.2.3 Focus on form ....................................................................................................................... 13
    2.2.4 Noticing ................................................................................................................................. 15
    2.2.5 Uptake .................................................................................................................................. 16
  2.3 Linguistic targets of corrective feedback .................................................................................... 17
  2.4 Implicit and explicit knowledge .................................................................................................. 19
  2.5 Role of working memory in corrective feedback ......................................................................... 20
  2.6 Contexts ..................................................................................................................................... 22
  2.7 Summary ..................................................................................................................................... 23
4.5.4 Factor analysis of implicit and explicit knowledge scores .............................................. 99
4.5.5 Analysis of working memory .......................................................................................... 100
4.5.6 Exit questionnaire ......................................................................................................... 101
4.5.7 Statistical procedures for data analysis ......................................................................... 101

CHAPTER 5  EFFECT OF RECASTS ON IMPLICIT KNOWLEDGE: RESULTS AND DISCUSSION ..........106
5.1 Statistical treatment .......................................................................................................... 106
5.2 Results ................................................................................................................................ 107
5.2.1 Results for 3rd person -s: Elicited imitation .................................................................. 107
5.2.2 Results for 3rd person -s: Oral production ................................................................. 112
5.2.3 Results for embedded questions: Elicited imitation ....................................................... 117
5.2.4 Results for embedded questions: Oral production ....................................................... 122
5.2.5 Results for reported noticing ......................................................................................... 127
5.3 Discussion .......................................................................................................................... 128
5.4 Conclusion ......................................................................................................................... 132

CHAPTER 6  EFFECT OF RECASTS ON EXPLICIT KNOWLEDGE: RESULTS AND DISCUSSION ..........133
6.1 Statistical treatment .......................................................................................................... 133
6.2 Results ................................................................................................................................ 134
6.2.1 Results for 3rd person -s: Written production .............................................................. 134
6.2.2 Results for 3rd person -s: Untimed grammaticality judgment ......................................... 139
6.2.3 Results for 3rd person -s: Untimed grammaticality judgment (ungrammatical items) ........ 145
6.2.4 Results for embedded questions: Written production ................................................... 152
6.2.5 Results for embedded questions: Untimed grammaticality judgment.............................. 156
6.2.6 Results for embedded questions: Untimed grammaticality judgment (ungrammatical items) .................................................................................................................. 160
6.3 Discussion .......................................................................................................................... 164
6.4 Summary ............................................................................................................................ 168

CHAPTER 7  UPTAKE AND ACQUISITION: RESULTS AND DISCUSSION ............................................169
7.1 Statistical treatment .......................................................................................................... 169
7.2 Results for uptake-with-repair .......................................................................................... 170
7.2.1 Results for repair frequency: 3rd person -s .................................................................. 170
7.2.2 Results for repair frequency: Embedded questions ....................................................... 171
List of Tables

Table 1: A Taxonomy of Corrective Feedback................................................................. 8
Table 2: Observational Studies of Recasts and Uptake .................................................. 27
Table 3: Studies of the Relationship between Recasts and Acquisition ...................... 36
Table 4: Comparative Studies of Recasting, Modelling vs Prompts in Laboratory Settings ...... 41
Table 5: Comparative Studies of Recasts and Prompts in Classroom Settings ............... 46
Table 6: Meta-Analyses of CF Involving Recasts .......................................................... 51
Table 7: My Partner's Life Style .................................................................................... 67
Table 8: Facts of Beijing/Shanghai/Guangzhou/Xi'an/Hangzhou/Lhasa/Shijiazhuang .......... 68
Table 9: Interview Transcript with Missing Questions .................................................. 70
Table 10: My Plan for an Eating Habits Survey ............................................................ 72
Table 11: Questions We Want to Ask Most .................................................................... 73
Table 12: EI Items for Practice ..................................................................................... 79
Table 13: EI Stimuli for Study 1 ...................................................................................... 80
Table 14: EI Stimuli for Study 2 ...................................................................................... 81
Table 15: OP Passage for Study 1 .................................................................................... 82
Table 16: OP Passage for Study 2 .................................................................................... 83
Table 17: WP Test for Study 1 ....................................................................................... 84
Table 18: WP Test for Study 2 ....................................................................................... 85
Table 19: Typical Items in UGJ Test .............................................................................. 87
Table 20: Nonword Sequences for Practice ................................................................. 89
Table 21: Nonword Sequences for the Test ................................................................... 90
Table 22: Strings of Random Numbers for the Test ..................................................... 92
Table 23: A Set of Sentences for Listening Span Practice ............................................ 94
Table 24: A Set of Sentences for Listening Span Test ................................................... 94
Table 25: Exit Questionnaire (English Version) ............................................................. 96
Table 26: Factor Loadings and Communalities: A One Factor Solution ....................... 99
Table 27: Factor Loadings and Communalities: A Two Factor Solution ....................... 100
Table 28: A Summary of Statistical Procedures for Data Analysis .............................. 104
Table 29: Descriptive Statistics of Scores for 3rd Person -s in EI Tests ......................... 108
Table 30: Adjusted Means for 3rd Person -s in EI Posttests .......................................... 110
Table 31: A Summary of within- and between-Subject Differences in EI Tests for 3rd Person -s ...... 111
Table 32: Descriptive Statistics of Scores for 3rd Person -s in OP Tests ......................... 112
Table 33: Adjusted Means for 3rd Person -s in OP Posttests ............................................................... 115
Table 34: A Summary of within- and between-Subject Differences in OP Tests for 3rd Person -s .... 116
Table 35: Descriptive Statistics of Scores for Embedded Questions in EI Tests ........................................ 118
Table 36: Adjusted Means for Embedded Questions in EI Posttests .................................................. 120
Table 37: A Summary of within- and between-Subject Differences in EI Tests for Embedded Questions ............................................................................................................................................ 121
Table 38: Descriptive Statistics of Scores for Embedded Questions in OP Tests .................................. 123
Table 39: A Summary of within- and between-Subject Differences in OP Tests for Embedded Questions ............................................................................................................................................ 126
Table 40: Descriptive Statistics for Reported Noticing of Explicit and Implicit Recasts ...................... 127
Table 41: Adjusted Means for 3rd Person -s in WP Posttests .............................................................. 135
Table 42: A Summary of within- and between-Subject Differences in WP Tests for 3rd Person -s .... 138
Table 43: Descriptive Statistics of UGJ Total Scores for 3rd Person -s ................................................. 139
Table 44: Adjusted Means for 3rd Person -s in UGJ Posttests ............................................................. 143
Table 45: A Summary of within- and between-Subject Differences in UGJ Tests for 3rd Person -s .... 144
Table 46: Descriptive Statistics of UGJ (Ungrammatical Item) Scores for 3rd Person -s ................... 146
Table 47: Adjusted Means for 3rd Person -s in UGJ Posttest 1 (Ungrammatical Item) ....................... 149
Table 48: Adjusted Means for 3rd Person -s in UGJ Posttest 2 (Ungrammatical Item) ....................... 149
Table 49: A Summary of within- and between-Subject Differences in UGJ (Ungrammatical Item) Tests for 3rd Person -s ................................................................................................................................... 151
Table 50: Descriptive Statistics of WP Scores for Embedded Questions ............................................ 153
Table 51: A Summary of within- and between-Subject Differences in WP Tests for Embedded Questions ............................................................................................................................................ 155
Table 52: Descriptive Statistics of UGJ Scores for Embedded Questions ........................................... 156
Table 53: Adjusted Means for Embedded Questions in UGJ Posttests .............................................. 158
Table 54: A Summary of within- and between-Subject Differences in UGJ Tests for Embedded Questions ............................................................................................................................................ 159
Table 55: Descriptive Statistics of UGJ (Ungrammatical Item) Scores for Embedded Questions ....... 161
Table 56: A Summary of within- and between-Subject Differences in UGJ (Ungrammatical Item) Tests for Embedded Questions .................................................................................................................... 163
Table 57: Uptake-with-Repair Frequencies for 3rd Person -s .............................................................. 170
Table 58: Uptake-with-Repair Frequencies for Embedded Questions ............................................... 171
Table 59: Repairers and Non-Repairers for 3rd Person -s .................................................................... 171
Table 61: Repairers and Non-Repairers for Embedded Questions ..................................................... 172
Table 62: The CR Group’s Repair and Gain Scores for 3rd Person -s ................................................... 173
Table 63: Correlations for the CR Group’s Repair and Gain Scores for 3rd Person -s ............................. 173
Table 64: The CR group’s Repair and Gain Scores for Embedded Questions .................................... 174
Table 65: Correlations for the CR Group’s Repair and Gain Scores for Embedded Questions ................. 174
Table 66: The IR Repair and Non-Repair Subgroups’ Gain Scores for 3rd Person -s ............................ 175
Table 67: T-tests of the IR Repair and Non-Repair Subgroups’ Gain Scores for 3rd Person -s ............... 175
Table 68: The CR Group’s Incorporation and Gain Scores for 3rd Person -s ...................................... 177
Table 69: Correlations for the CR Group’s Incorporation and Gain Scores for 3rd Person -s ............... 177
Table 70: The CR Group’s Incorporation and Gain Scores for Embedded Questions ......................... 178
Table 71: Correlations for the CR Group’s Incorporation and Gain Scores for Embedded Questions 178
Table 72: The IR Incorporation and Non-Incorporation Subgroups’ Gain Scores for 3rd Person -s ....... 179
Table 73: T-tests of the IR Incorporators and Non-Incorporators’ Gain Scores for 3rd Person -s ............ 179
Table 74: The CR Group’s WM Capacity and Gain Scores for 3rd Person -s ........................................ 189
Table 75: The CR Group’s WM-Gain Correlation Coefficients for 3rd Person -s .................................. 190
Table 76: The CR Group’s WM Capacity and Gain Scores for Embedded Questions .......................... 190
Table 77: The CR Group’s WM-Gain Correlation Coefficients for Embedded Questions ....................... 191
Table 78: The IR Group’s WM Capacity and Gain Scores for 3rd Person -s ........................................ 191
Table 79: The IR Group’s WM-Gain Correlation Coefficients for 3rd Person -s ................................... 192
Table 80: The IR Group’s WM Capacity and Gain Scores for Embedded Questions .......................... 193
Table 81: The IR Group’s WM-Gain Correlation Coefficients for Embedded Questions ....................... 193
List of Figures

Figure 1: Research Design ..................................................................................................................... 59
Figure 2: Celebrity Pictures ................................................................................................................... 65
Figure 3: Means for 3rd Person -s in EI Tests over Time by Group ..................................................... 108
Figure 4: Means for 3rd Person -s in OP Tests over Time by Group ................................................... 113
Figure 5: Means for Embedded Questions in EI Tests over Time by Group ..................................... 118
Figure 6: Means for Embedded Questions in OP Tests over Time by Group ................................... 123
Figure 7: Means for 3rd Person -s in WP Tests over Time by Group .................................................. 135
Figure 8: Means for 3rd Person -s in UGJ Tests over Time by Group ............................................... 140
Figure 9: Means for 3rd Person -s in UGJ (Ungrammatical Item) Tests over Time by Group .......... 146
Figure 10: Means for Embedded Questions in WP Tests over Time by Group .............................. 153
Figure 11: Means for Embedded Questions in UGJ Tests over Time by Group .............................. 157
Figure 12: Means for Embedded Questions in UGJ (Ungrammatical Item) Tests over Time by Group ................................................................................................................................. 161
## Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANCOVA</td>
<td>analysis of covariance</td>
</tr>
<tr>
<td>ANOVA</td>
<td>analysis of variance</td>
</tr>
<tr>
<td>CF</td>
<td>corrective feedback</td>
</tr>
<tr>
<td>CR</td>
<td>corrective recasts</td>
</tr>
<tr>
<td>CWM</td>
<td>complex working memory</td>
</tr>
<tr>
<td>EFL</td>
<td>English as a foreign language</td>
</tr>
<tr>
<td>EI</td>
<td>elicited imitation</td>
</tr>
<tr>
<td>ESL</td>
<td>English as a second language</td>
</tr>
<tr>
<td>IR</td>
<td>implicit recasts</td>
</tr>
<tr>
<td>L2</td>
<td>second language</td>
</tr>
<tr>
<td>NNS</td>
<td>non-native speaker</td>
</tr>
<tr>
<td>NS</td>
<td>native speaker</td>
</tr>
<tr>
<td>OP</td>
<td>oral production</td>
</tr>
<tr>
<td>PSTM</td>
<td>phonological short-term memory</td>
</tr>
<tr>
<td>SLA</td>
<td>second language acquisition</td>
</tr>
<tr>
<td>TkC</td>
<td>task control</td>
</tr>
<tr>
<td>TtC</td>
<td>test control</td>
</tr>
<tr>
<td>UGJ</td>
<td>untimed grammaticality judgment test</td>
</tr>
<tr>
<td>WM</td>
<td>working memory</td>
</tr>
<tr>
<td>WP</td>
<td>written production</td>
</tr>
</tbody>
</table>
Chapter 1 Introduction

When an L2 learner commits an error, the teacher or the competent interlocutor may correct it, indicating or concealing the nature of the error, provide clues for the learner to self-correct, or combine any of these techniques. This corrective feedback (CF) on a learner’s error has attracted considerable interest from SLA researchers for decades for theoretical and pedagogical reasons. The study of CF creates an interface between theory and practice (Sheen, 2011). The most contentious oral CF is recasting, which refers to an error-free reformulation of the learner’s nontargetlike utterance keeping the meaning intact. Descriptive studies in a classroom setting show that recasts are the preferred form of correcting errors among language teachers but lead to fewer instances of uptake-with-repair than other types of CF (e.g. Lyster & Ranta, 1997; Panova & Lyster, 2002; cf Ellis, Basturkmen & Loewen, 2001). However, experimental studies in a laboratory setting reveal that recasts are generally effective (Long, Inagaki, & Ortega, 1998; Lyster & Izquierdo, 2009; Mackey & Philp, 1998; McDonough & Mackey, 2006). Quasi-experimental studies in a classroom context have also examined the role of recasts in instructed SLA, producing mixed results (e.g. Doughty & Varela, 1998; Goo, 2012; Révész, 2012; cf Lyster, 2004; R. Ellis, 2007).

The study reported in this thesis belongs to the last line of research. It conducted an experiment (4 groups by 3 times of test) in an EFL classroom context to investigate the effect of recasts by considering the explicitness of the CF, the salience of the target structure, measures of linguistic knowledge, uptake of the CF and the role of individual differences in working memory (WM).

1.1 Aim of the study

This study aims to explore the mechanism of recasts by examining the interactions between CF, target structure, tests, uptake and WM. Explicit and implicit recasts were manipulated to find out whether explicitness makes a difference to the efficacy of recasts. This sheds light on the current debate about whether recasts contribute to acquisition through providing positive and negative evidence or positive evidence only. The feedback targeted 3rd person singular -s and embedded questions, which are different in terms of saliency (more relevant for implicit knowledge) and complexity (more relevant for explicit knowledge), so the comparison of these two structures helps to clarify whether the choice of linguistic feature influences the effect of recasts on acquisition. The effectiveness of recasts was measured using implicit and
explicit knowledge tests so as to examine on which type of knowledge recasts have an effect. Correlations of gain scores for implicit or explicit knowledge and uptake rate were calculated to investigate whether the uptake following recasts was related to acquisition. Correlational analyses of gain scores and WM capacity were computed to examine whether individual differences in WM mediate the effectiveness of recasts. To sum up, this study examines the effects of explicit and implicit recasts on the development of implicit and explicit knowledge of two different grammatical structures and the roles of uptake and WM in the recasts-driven acquisition in order to understand under what conditions recasts facilitate L2 acquisition.

1.2 Theoretical motivation

This study was motivated by theoretical positions on the role of corrective feedback. Research on CF has been informed by a number of questions: 1). Does negative evidence contribute to L2 acquisition in addition to positive evidence? 2). Does CF facilitate the development of implicit knowledge as well as explicit knowledge? 3). Does didactic CF facilitate acquisition in the same way as interactional CF? 4). Does CF trigger noticing or noticing the gap? 5). Is uptake of CF important for acquisition? 6). What factors mediate the role of recasts in L2 acquisition?

Nativists posit that L2 is acquired like L1, by the speaker being exposed to positive evidence (information about what is acceptable in the target language). For instance, Krashen (1981, 1982) claimed that comprehensible input is sufficient for acquisition. For him, negative evidence (information about what is not acceptable) is likely to arouse anxiety, interfere with comprehension and thus be detrimental (1994). In contrast, cognitive-interactionists argue that negative evidence is not necessary but does facilitate L2 acquisition. Negotiation of meaning including CF integrates the advantages of input, output and selective attention (Long, 1996). Krashen (2003) also mitigated acknowledging that CF might facilitate learning.

Some UG-based researchers (e.g. Schwartz, 1993) have argued that corrective feedback contributes to learned knowledge (explicit knowledge) but not to linguistic competence (implicit knowledge). Cognitive-interactionists, however, attribute a pivotal role to CF in L2 acquisition. CF arising from the negotiation for meaning serves to draw learners’ attention to form, leading to form-meaning mapping. In order to investigate these different positions, R. Ellis (2004, 2005) argued that separate tests for implicit knowledge and explicit knowledge are needed.
It should be noted that the cognitive-interactionists emphasized the value of implicit CF arising from the negotiation for meaning because it encourages joint attention to form and meaning. They argued that explicit CF creates a “time-out” from communication and treats language as an object. However, as Lyster (2001, 2002) noted, didactic CF (negotiation of form) can also drive acquisition as long as it occurs in a communicative context because it pushes learners to retrieve the form from memory and this pushed output proceduralizes and automatizes linguistic knowledge.

Long (1996, 2007) and Doughty (2001) assumed that although recasts are implicit, they are likely to draw the learner’s attention to form because what is negotiated is the learner’s message, and thus they assist a cognitive comparison between what the learner has said and the recast. Noticing and noticing the gap are crucial mechanisms to induce the restructuring of interlanguage.Recasts provide not only positive evidence but also negative evidence, which in turn facilitates acquisition. In contrast, Lyster (1998b) claimed that recasts are ambiguous as to whether they are corrective or merely functional as confirmation checks and that the corrective function is too implicit to be perceived.

Uptake refers to learners’ responses towards corrective feedback. Lyster and colleagues took uptake as an important indicator of the effectiveness of CF. They drew on the Skill-Building Theory (DeKeyser, 2001, 2007) and the Output Hypothesis (Swain, 1985, 2000) to claim that prompts lead to more uptake, which in turn strengthens the control of linguistic knowledge, so they are more effective than recasts. Uptake is of importance for acquisition also because it indicates noticing of the corrective intention. However, lack of uptake does not necessarily mean a lack of noticing. It is possible that learners noticed the correction but produced no uptake either because uptake of recasts is optional or because there is no opportunity for them to do so when the topic continues. Uptake cannot be equated with acquisition and it remains unclear whether uptake of recasts is related to acquisition.

The efficacy of CF may be mediated by cognitive factors, such as the learner’s working memory capacity. WM is regarded as a component of language aptitude (Miyake & Friedman, 1998; Skehan, 2002) and implicated in language comprehension (e.g. Harrington & Sawyer, 1992; Leeser, 2007), vocabulary learning (e.g. Cheung, 1996) and grammatical learning (e.g. N. Ellis, 2005; Juffs, 2004; Sagarra, 2007). However, the role of WM in processing CF including recasts is still under-researched. The few studies that have investigated the
relationship between WM and recasts have obtained mixed results. For example, Li (2013b) and Yilmaz (2013) reported no relationship.

The effectiveness of CF seems to vary depending on the target structure. As R. Ellis (2006a) argued, the learning difficulty of different linguistic features may affect the role that CF plays in L2 acquisition. Research on CF comparing different structures is interesting but rare to date.

1.3 Practical concerns

As an EFL teacher, I have long been puzzled about error correction. For instance, should we correct students’ errors? What errors should we correct: pronunciation, vocabulary, grammar, pragmatic strategies, or all of them? Should we focus on one target structure intensively or correct any error that we encounter? Should we correct directly or indirectly without indicating the nature of the error? Should we use CF strategies consistently or arbitrarily? When should we correct, immediately or when a task has been completed? How frequently should we correct, every time we hear or see an error, or selectively? I think that most teachers handle CF intuitively but often find that it does not work efficiently. We often hear language teachers complaining that their students commit the same error after receiving CF.

The history of language teaching has witnessed pendulum swings in attitudes towards CF (Pawlak, 2014). The Audiolingual Method, based on behaviourism, involved pattern practice and, where necessary, corrective feedback to prevent learners from forming bad habits. In contrast, the Natural Approach proposes that instruction consist of a series of communicative activities providing comprehensible input and proscribes the use of CF to foster the low-anxiety atmosphere claimed to be necessary for acquisition. Presentation-Practice-Production (PPP) – probably the most popular approach as reflected in mainstream text books – acknowledges the need for corrective feedback at least during the ‘practice’ stage. Harmer (2001, p.105) distinguished “communicative” from “non-communicative” activities and stated that feedback is not compatible with the former, only with the latter. However, Task-Based Instruction (a strong version of the Communicative Language Teaching Approach), which draws on SLA theories and research, views CF as essential for attracting attention to form while learners are communicating (R. Ellis, 2005a).

As discussed in the previous section, Long (1991) advocated focus on form as an ideal type of instruction, where the overriding focus is on meaning, and negotiation for meaning
involving recasts induces attention to form. However, Lyster (1998b, 2004) supported negotiation of form and the use of prompts as he claimed that they are more effective than recasts. Controversy surrounding recasts in classroom settings remains, especially when and how they drive instructed SLA. It is unclear whether recasts work for particular types of linguistic features and language knowledge. It also remains unclear what kinds of learners benefit from recasts, and whether explicit recasts are more effective than implicit ones. Therefore, it is of practical significance to research the interactions between the characteristics of recasts, target features, measures of linguistic knowledge and individual differences in task-based lessons.

1.4 Structure of the thesis

This thesis consists of nine chapters. Chapter 1 introduces the background of the study reported in the thesis. It presents the overall aim of the study, explains its theoretical motivation and practical concerns, and outlines the organization of the whole thesis. Chapter 2 defines the key terms and, reviews theoretical positions on the role of corrective feedback in L2 acquisition, as well as the controversies surrounding recasts in relation to target features, linguistic knowledge, individual differences in working memory and learning contexts. Chapter 3 reviews empirical studies on the efficacy of recasts in terms of uptake, noticing and acquisition. It discusses three types of comparative research: classroom-based descriptive studies, laboratory-based experimental studies and classroom-based quasi-experimental studies. This chapter also considers research that has examined whether or not working memory capacity mediates the effect of recasts. Chapter 4 describes the methodology of the study, including research questions, research design, participants, instructional tasks, treatment (explicit vs implicit recasts), language tests (implicit vs explicit knowledge), working memory tests, scoring procedures and statistical procedures. Chapter 5 reports and discusses the results for the effects of corrective recasts and implicit recasts on the development of implicit knowledge. Chapter 6 reports and discusses the effects of the two types of recasts on the development of explicit knowledge. Chapter 7 uses the results of correlational analyses to address the relationship between uptake following recasts and subsequent acquisition as implicit and explicit knowledge. Chapter 8 employs the results of correlational analyses to investigate the role of phonological short-term memory and complex working memory in mediating the effect that recasts have on acquisition. The WM-uptake relationship and WM-gain relationship are both examined. Chapter 9 concludes with the main
findings, theoretical implications, pedagogical implications and limitations of this study, and suggestions for future research.
Chapter 2 Theoretical Issues

The study reported in this thesis investigates the role of a particular type of corrective feedback (CF), recasts, in the development of implicit and explicit knowledge and the relationship between uptake following recasts, language development, and individual differences in working memory (WM). The general purpose is to find out in what conditions, and how, recasts facilitate L2 acquisition. To provide a rationale for this research, Chapter 2 begins by defining CF and the type of CF of particular interest (i.e. recasts), and introducing theoretical positions relating to CF. Then, it looks at the efficacy of recasts from the perspective of linguistic targets. After that, it defines implicit and explicit knowledge in order to see which type of knowledge recasts potentially help to develop. It also considers the mediation of working memory (WM) capacity on the effectiveness of the CF. Finally, this chapter briefly discusses in what way context contributes to the beneficial effect of recasts.

2.1 Corrective feedback

Corrective feedback has been used interchangeably with the following terms: “negative evidence” by linguists, “repair” by discourse analysts, “negative feedback” by psychologists, and “focus-on-form” by cognitive interactionists (Gass, 1997; Lyster & Ranta, 1997). Second language teachers prefer to use “corrective feedback” to address the teacher’s reaction to a learner’s error, in writing or speaking. What this thesis focuses on is oral corrective feedback, that is, the teacher’s verbal responses towards a learner’s utterance containing an error or errors. Such a response may indicate the error only, provide the correct linguistic form, explain the nature of the error, or employ any combination of these devices (Ellis, Loewen, & Erlam, 2006).

2.1.1 Types of corrective feedback

Second language (L2) teachers have at their disposal different strategies to correct a learner’s error. Lyster and Ranta (1997) identified from their immersion class transcripts six types of corrective feedback, namely, explicit correction (explicit provision of the correct form), recast (error-free reformulation of the erroneous utterance), clarification request (indication of misunderstanding or ill-formedness), metalinguistic feedback (comments on ill- or well-formedness), elicitation (eliciting the correct form) and repetition (isolated repetition of the error). Lyster (1998b, 2001, 2004) re-categorized the last four interactional moves as prompts as they are all more likely to push learners to self-correct by withholding the correct form.
Carroll and Swain (1993), Ellis et al. (2006) and other researchers distinguish types of CF in terms of their implicitness. Recasts are usually categorized as implicit CF as they do not overtly indicate that there is an error, whereas explicit correction and metalinguistic feedback are considered to be explicit.

The above two distinctions are both of theoretical significance. Cognitive interactionists argued for input-providing CF, considering input as the driving force of acquisition while skill-building theories favored output-prompting CF, regarding output as a key mechanism for proceduralization and automatization. As for the implicit-explicit distinction, Long (1996) proposed that implicit CF is an ideal form for acquisition as it does not interrupt the communicative flow and thus induces form-function mapping. However, explicit forms of correction are assumed to be more effective than implicit ones according to the Noticing Hypothesis (Schmidt, 1994; 2001). It is likely that explicit CF leads to explicit knowledge only (Krashen, 1985). Whether it contributes to acquisition of implicit knowledge is controversial. R. Ellis (2009) combined these two separate distinctions into a comprehensive taxonomy of CF (as shown in Table 1). This framework systematically accommodates different types of CF but, as R. Ellis acknowledged, fails to reflect the variation of a single type of CF. He took recasts as an example. Although recasts are described as “input-providing, implicit” in the framework, they in fact vary in implicitness, for instance, depending on whether they are combined with other types of CF, whether there is emphasis on the error, whether they involve declarative or interrogative intonation, and whether they occur in full or partial form, etc.

### Table 1: A Taxonomy of Corrective Feedback

<table>
<thead>
<tr>
<th></th>
<th>Implicit</th>
<th>Explicit</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Input-providing</strong></td>
<td>Recast</td>
<td>Explicit correction</td>
</tr>
<tr>
<td><strong>Output-prompting</strong></td>
<td>Repetition</td>
<td>Metalinguistic explanation</td>
</tr>
<tr>
<td></td>
<td>Clarification request</td>
<td>Elicitation</td>
</tr>
</tbody>
</table>

Note. This table is adapted from R. Ellis’ (2009, p.8) study, omitting “paralinguistic signal” as this research only addresses verbal CF.

#### 2.1.2 Types of recasts

Recasts in this research refer to the teacher’s targetlike reformulation of a learner’s utterance
containing a grammatical error in a task-based environment, keeping the original meaning intact. Previous studies of recasts manipulated and defined recasts so differently that it was impossible to generalize their results because they “were not looking at the same thing” (Ellis & Sheen, 2006, p.578). For instance, Doughty and Varela’s (1998) recasts are made up of repetition of the error, emphasis on the error and a reformulation. In contrast, Lyster (1998b, 2002, 2004) operationalized the CF as an error-free reformulation. What renders Lyster’s CF more implicit is that the CF is accompanied with signals such as “oui [yes]” or “c’est bien [very good]” and sometimes is incorporated into a long utterance, indistinguishable from non-corrective repetition. Unlike Lyster, Long took into consideration only conversational recasts initiated by a communicative breakdown as shown in his definition: “a reformulation … the focus of the interlocutors is on meaning, not language as object” (2007, p.77). Arguably, conversational recasts are more implicit than didactic recasts (Lyster, Saito, & Sato, 2013). It is evident that recasts are not a homogeneous construct but take different forms ranged along an implicit-explicit continuum (Ellis & Sheen, 2006).

This research compares the effects of explicit and implicit recasts. Following Doughty and Varela’s (1998) example, an explicit recast was realized by a repetition to draw attention followed by a recast to provide the contrastive L2 form. An example of the explicit recasts the current research employed is presented below.

S27: He often say “pia pia di”.
T: He often SAY?
S27: Say.
T: Or he often SAYS.
S27: Er, says, says.

When Student No. 27 committed a 3rd person -s error, the teacher first repeated a part of the erroneous utterance, “He often SAY?”, emphasizing the error “SAY”. The student failed to recognize the teacher’s corrective intention or was unable to produce the correct form. Then, the teacher recast the erroneous part of the learner’s utterance. When an exemplar of the target structure was provided in a recast, it was also emphasized. This supra-segmental feature is “potentially a variable” as it may “influence the extent to which the corrective force is explicit to learners” (Sheen, 2006, p. 364). The student’s uptake “says, says” indicates that
she attended not only to the form “-s” but also to the corrective force of the recast.

Implicit recasts only involve a “reformulation of a learner’s nontargetlike utterance into a targetlike one” (Nassaji, 2009, p.412). An implicit recast is operationalized using Lyster’s (2004) model, which consists of a turn-taking indicator and a single recast of the ungrammatical utterance. An example from the current study is given below. Hearing the recast, the student did not repair her error, but carried on with her topic. The overlapping turns and the failure to self-correct suggest that the student may have failed to notice the correction.

S45: In a movie, he pretend to be a, a blind person.
T:     Hmm. In a movie, he usually plays, pretends = to be.
S45: = pretend to be a blind person and a beauty, and beauty wal, wal, walk beside her, she said, “whose wallet?”

As this example illustrates, the distinction between implicit and explicit recasts can be verified by observing learners’ reaction to the recast (i.e. whether they uptake the corrective move and if so how). Uptake is a key issue because it can indicate whether the learner has perceived the corrective intention of a recast. If the learner does so, the CF provides negative evidence.

2.1.3 Positive and negative evidence

Long (1996) classified environmental input for language learning into two types: positive and negative evidence. Positive evidence refers to the input that learners receive as to what is grammatical or acceptable in the target language. It is provided as natural input in authentic communication or as modified input in teacher talk. Positive evidence in a classroom setting may take the form of naturally occurring interaction or pre-selected instructional materials. It is impossible to acquire a language without being exposed to the correct exemplars of the language. Krashen (1981, 1985, 1994) claimed that it is the comprehensible positive evidence which is a bit beyond the current level of competence (i.e. “i+1”) that drives L2 acquisition. There is a wide consensus among theorists that positive evidence is necessary (Gass, 2003); what constitutes an issue of controversy is the role of negative evidence.

Negative evidence is the information that learners receive concerning what is ungrammatical
or unacceptable in the target language. It can be provided in the form of grammar instruction preemptively or in the form of corrective feedback reactively. There are different theoretical positions on the role of negative evidence in the process of L2 acquisition, which will be discussed in the following section. This section focuses on whether recasts provide negative evidence as well as positive evidence, which is contentious in literature (Ellis & Sheen, 2006; Nicholas, Lightbown, & Spada, 2001). Doughty (2001), Long (1996, 2007), N. Ellis (1995) and Goo and Mackey (2013) posited that they do so whereas Lyster (1998b, 2004) argued that the corrective function of recasts is too implicit to be perceived. It is not easy to distinguish corrective reformulation via recasts from non-corrective repetition for confirmation or confirmation checks. Leeman (2003) empirically researched the effects of positive evidence (implicit recasts), negative evidence, and enhanced salience. Her conclusion was that recasts’ contribution to L2 learning lies in the fact that they offer positive evidence rather than negative evidence.

Ellis and Sheen (2006) proposed a solution to this controversy by looking at how learners interpret the illocutionary force of the feedback. If the learners report noticing of the correction or show signs of noticing in their uptake, the CF can be regarded as providing negative evidence. On the other hand, if they fail to perceive the correction, the CF may only constitute a source of positive evidence.

Corrective recasts, which are made up of a repetition, a recast and added supra-segmental features, are linguistically more salient to learners, and thus are more likely to be perceived as “corrective”. This type of recast is hypothesized to provide both positive and negative evidence. In contrast, implicit recasts comprised of a turn-taking signal and an error-free reformulation are assumed to only serve as positive evidence. If so, it can be argued that corrective recasts are more likely to facilitate language learning because they potentially provide negative evidence, which is facilitative of L2 acquisition according to cognitive-interactionist theories (e.g. Long, 1996).

### 2.2 Theoretical basis for corrective feedback

L2 acquisition theories differ on the utility of negative evidence, the comparative efficacy of different types of corrective feedback, and the acquisitional value of uptake following CF. Some theorists take a zero position on corrective feedback while other theorists allocate it a fundamental role.
2.2.1 Opponents of corrective feedback

Nativists assume that all human beings possess an innate mechanism for language learning. For instance, Chomsky’s (1975) UG grammar is a language acquisition device, whose parameters are set through exposure to natural input, that is, positive evidence. Some UG-based L2 researchers (e.g. Schwartz, 1993) distinguish “learned linguistic knowledge” and “competence” and assume that negative evidence only contributes to the development of the former but not the latter. What UG feeds on is positive input in both L1 and L2 acquisition. That is to say, these UG advocates maintain that language competence is built on the basis of positive evidence and that negative evidence is unnecessary.

Krashen (1981, 1982, 1985) posited that there is no interface between subconscious acquisition and conscious learning. Language knowledge gained from learning only helps monitor the production of acquired knowledge. For him, explicit instruction does not contribute to acquisition. All that is needed is comprehensible input (i.e. positive evidence). He argued that corrective feedback, or “error correction [including recasts] has little or no effect on subconscious acquisition” (1982, p.11), “is a serious mistake” (1982, p.74) and is even detrimental if it interrupts comprehension and causes anxiety in learners. He suggested that teachers should overlook student errors which will disappear over time as acquisition takes place. However, this zero position on CF conflicts with most recent L2 acquisition theories.

2.2.2 Proponents of corrective feedback

Cognitive-interactionists hypothesize that positive evidence is not sufficient for acquisition. Gass (1991) proposed that for input to be internalized learners have to notice the mismatch between the input and their own interlanguage. “Noticing” and “noticing the gap” (Schmidt & Frota, 1986) are psychological processes where corrective feedback is believed to be crucial. In Long’s (1996) revised Interaction Hypothesis, negotiation for meaning following a breakdown in communication triggers CF moves, which may draw learners’ attention to form. This micro-process in working memory leads to subsequent acquisition. Cognitive-interactionists posit that corrective feedback may not be necessary for acquisition but is facilitative when it occurs.

Pushing output is another way in which CF contributes to acquisition. As Lyster (2004) argued, corrective feedback that withholds the correct form elicits self-repair from the learner. Such “pushed output” may “trigger noticing”, act as “a way of hypothesis testing” and help
the learner to “reflect on metalinguistic information”, and also “enhance fluency” (Swain, 1995, pp.129-133).

These two ways of viewing the role of CF in acquisition differ in terms of which type of CF they claim is more effective. The former emphasizes input and proposes recasts as an ideal form of CF whereas the latter emphasizes output and the need for prompts.

### 2.2.3 Focus on form

Long (1991) distinguished between focus on forms and focus on form. Lessons with a focus on forms teach isolated linguistic forms as content, treating language as an object. On the other hand, a lesson with a focus on form “overtly draw[s] students’ attention to linguistic elements as they arise incidentally in lessons whose overriding focus is on meaning, or communication” (pp.45-46). Focus on form requires the following: “designing tasks to promote learner engagement with meaning prior to form; seeking to attain and document task essentialness or naturalness of the L2 forms; attempting to ensure that instruction was unobtrusive; [and] documenting learner mental processes (‘noticing’)” (Norris & Ortega, 2000, p.438). It should be noted that Long’s original focus on form has been stretched. As Ellis et al. (2001) pointed out, there are pedagogically two types of focus on form: preemptive and reactive. Preemptive focus on form draws explicit attention to a linguistic feature to avoid an error whereas reactive focus on form corrects a learner’s error during meaning communication to provide negative evidence.

Long’s focus on form is reactive and incidental. It occurs when there is a communicative breakdown. “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” (Long, 1996, pp.451-452). “Negative feedback obtained during negotiation work or elsewhere may be facilitative of L2 development” (p.414). In this revised version of the Interaction Hypothesis, negotiation of meaning not only increases comprehensibility but also triggers attention to form. Long’s negotiation of meaning includes input modification (e.g. stress on key words), semantically contingent responses (e.g. recasts) and conversational modifications (e.g. clarification requests). Among these interactional features, Long favors recasts, stating that the reformulation type of CF ensures the message is at least partially clear to the learner so that he/she can free up resources for form. Doughty also contends that an immediate contingent recast is most “promising” because it “can easily
fit into WM along with the original utterance to which it is to be compared”. “Such cognitive comparison does lead to form-function-meaning mapping” (2001, p. 257) and consequently results in interlanguage restructuring. Both Long and Doughty considered that recasts are implicit in nature, do not interrupt the communicative flow, and thus constitute an ideal form of feedback. For Long (2007), obtrusive CF, such as prompts (metalinguistic CF in particular) and explicit correction, may be detrimental to acquisition because they discourage joint attention to meaning and form. So far, there is no evidence to support this claim.

Lyster (2001) claimed that the term “negotiation of meaning” cannot precisely capture the characteristics of focus on form in the classroom. He proposed that “negotiation of form” is also needed. He contended that negotiation of form when there is no communicative breakdown is equally effective in promoting L2 development as long as it occurs in a meaning-focused context. Negotiation of form includes metalinguistic clues, elicitation, repetition and clarification requests (i.e. prompts). Lyster (2002) acknowledged the problematicity of the form-meaning dichotomous distinction. Recasts do not necessarily fall into the category of negotiation of meaning. When they are short with the correct form isolated or emphasized with stress, the corrective function of recasts is obvious. This type of recast may create an opportunity for the teacher and students to step out of communication temporarily and treat language as an object. The key difference between the negotiation of meaning and form actually lies in whether there is a communicative breakdown. Recasts in the current research were not motivated by a communicative problem but by the occurrence of a grammatical error, so they constitute a type of negotiation of form as prompts do. However, prompts such as metalinguistic clues, elicitation, repetition and clarification requests do not provide the correct form but give clues for learners to repair their utterances utilizing their own resources. Drawing on uptake (1998) and posttests (2004), Lyster supported the comparative efficacy of prompts over recasts, especially in learning linguistic features that had already been partially acquired.

However, it may be meaningless to compare the effects of prompts and recasts. Goo and Mackey (2013) called this a comparison of “apples and oranges” (p.148) because prompts and recasts are inherently different. Uptake following a prompt is obligatory in terms of discourse appropriateness whereas uptake of recasts is optional. Prompts and recasts potentially assist acquisition in very different ways as Ellis and Shintani (2014) noted. Prompts possibly increase the control over partially acquired forms but recasts assist the
internalization of new forms. They further suggested that teachers should combine these two types of CF, for instance, by repeating the error as a prompt and then recasting the erroneous utterance if no self-repair occurs. However, Lyster and Ranta (2013) insisted that “apples and oranges” should be compared precisely because they are inherently different. In practice, language teachers will benefit from the knowledge as to which type of CF is more facilitative, especially by becoming aware of what situation one works better in than another. L2 researchers remain interested in investigating this comparison.

2.2.4 Noticing

Whether the development of linguistic competence (implicit knowledge) involves consciousness has been the focus of considerable research. Krashen’s (1981, 1982) construct of acquisition is unconscious. However, Schmidt (1990, 1993) stated that noticing (the conscious registration of the linguistic forms in input) is a necessary and sufficient condition for acquisition in his original Noticing Hypothesis. He acknowledged incidental learning without intention and implicit learning without understanding rules, but doubted subliminal learning without awareness. He argued that learning cannot take place without conscious attention to linguistic forms. In his revised version of the Noticing Hypothesis (1994, 1995; 2001), Schmidt modified his position slightly by claiming that implicit learning (i.e. learning without noticing) may occur but that the more learners notice the more they learn. This statement needs empirical evidence to support. R. Ellis (1997) also described the process of learning implicit knowledge as involving “noticing”, “comparing” and “integrating” (p.119). In his model of language acquisition, noticing works as a gateway through which input enters working memory, that is to say, he posited that only what learners have noticed in the input becomes intake for further analysis. Proponents of noticing also attribute a role to “noticing the gap” (Schmidt & Frota, 1986), where there is conscious awareness of a mismatch between the correct form in input and the error by the learner. This type of cognitive comparison in working memory possibly triggers interlanguage restructuring in long-term memory.

Schmidt (1990) identified instruction, frequency, perceptual salience, skill level and task demands as factors that affect noticing in input. Corrective recasts are more likely to be perceived and the target structure is more likely to be salient to learners. From this perspective corrective recasts will facilitate language acquisition more efficiently. However, empirical studies are needed to investigate whether the nature of recasts affects the extent to
which attention to form and acquisition takes place.

Truscott (1998) doubted the theoretical foundations and the validity of Schmidt’s noticing hypothesis. He contended that conscious noticing may be tied only to the acquisition of metalinguistic knowledge. To confirm whether noticing is necessary for the development of L2 competence, appropriate measures of implicit knowledge are needed. Answering this call, the current research obtained not only measures of explicit knowledge but also those of implicit knowledge.

2.2.5 Uptake

Uptake is another object of inquiry in CF research. It refers to a learner’s responses to corrective feedback. Lyster and Ranta defined it 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” (1997, p. 49). They further classified uptake into two types: “repair” and “needs repair”.

Whether repair following CF contributes to language learning remains contentious. Ellis et al. (2001) claimed that uptake with repair is facilitative of acquisition. Lyster (1998; 2004) also posited that uptake of CF is essential for acquisition by referring to the skill-learning theory. The theory states that acquisition of a language, like learning other skills, goes through three stages: declarative, procedural and automatic (DeKeyser, 1998, 2001, 2003, 2007). Transition from declarative to procedural knowledge and then from procedural to automatic knowledge is realized via repeated practice and feedback. Prompts create opportunities for pushed output (i.e. uptake). To produce self-repair, learners have to retrieve knowledge about a partially acquired structure from their own memory. The retrieval of more targetlike forms is automatized via pushed output. Thereby, output-pushing CF assists learners to gain control over their partially acquired structures. In this sense, prompts are more likely to facilitate acquisition than recasts because the latter does not require repair on the part of the learner and in fact often repair does not occur.

In contrast, uptake does not have a substantial place in Long’s Interaction Hypothesis (1991, 1996). Long emphasized the efficacy of the implicit input-providing type of CF arising from negotiation for meaning drawing on the Comprehensible Input Hypothesis, the Noticing Hypothesis and theories on saliency (e.g. input enhancement). For him, recasts do not interrupt the communicative flow due to their implicitness, which is an advantage because it
fosters joint attention to both form and meaning, inducing form-function mapping (1991, 2007). The reformulation provided via recasts makes the intended meaning clear to the learner, so he can spare more attentional resources to the form targeted by the CF (1996; 2007). The learner may be motivated to attend to the linguistic change in a recast because what is at stake is his message. Recasts are contingent on the learner’s error and this contingency suggests the juxtaposition of the learner’s deviant form and the teacher’s correct form, which provides an ideal opportunity for cognitive comparison between input and output. The observed gap may finally lead to restructuring in interlanguage. Therefore, interactional modification involving recasts integrates the input, learner internal capacity and output into an effective mechanism for acquisition (2007). Long argued that the benefits of recasts should be evaluated by means of posttests instead of immediate uptake because their effect may not be immediately apparent. “Immediate incorporation of the feedback in a learner’s next utterance is an unreliable measure of effectiveness” (2007, p.94) given the discourse constraints, especially in the classroom. Mackey and Philp (1998) argued that responses to recasts do not predict the effect of the CF on language development and they are probably “red herrings” (p. 338).

Lyster et al. (2013), however, reiterated the claim that uptake of prompts is an indicator of learning outcome whereas repetition-type repair following recasts is not. Corrective recasts in the current research are composed of a repetition and a recast, so they are more likely to elicit uptake than implicit recasts, which involve only a single recasting move. To compare corrective and implicit recasts at the levels of uptake and posttest helps to elucidate the role of uptake following recasts. Hearing recasts, learners may simply repeat the teacher’s CF or incorporate the provided correct form into a longer utterance if they perceive the corrective force and choose to repair. Also of theoretical significance is whether repetition and incorporation are differently associated with L2 learning.

### 2.3 Linguistic targets of corrective feedback

The effectiveness of corrective feedback seems to vary depending on the nature of the linguistic feature it targets, as shown in uptake rate. Lyster et al. (2013) noted that there is more uptake of CF on lexical and pronunciation errors than morphosyntactic errors reflecting the importance of the former for communication. The more the target features tend to affect communication, the more amenable to CF they are. However, empirical research on this issue is clearly needed.
As for different grammatical targets, Ellis (2007) suggested that grammatical difficulty can account for the different effects of CF reported in previous studies. Referring to Pienemann’s (1998) Processibility Theory, he argued that the structures that many previous studies investigated were beyond learners’ ability to acquire as implicit knowledge. He also noted that target structures differ in terms of how easy or difficult they are to learn as both implicit and explicit knowledge. Ellis (2006a, p.435) proposed “frequency”, “saliency”, “functional value”, “regularity” and “processability” as the criteria for evaluating the learning difficulty of different linguistic features as implicit knowledge and “conceptual clarity” and “metalanguage” as the criteria for explicit knowledge. For example, 3rd person -s is phonologically non-salient, lacking in communicative value, and in terms of Processibility Theory involves information exchange between sentence constituents (i.e. an advanced processing operation), so it is difficult to acquire as implicit knowledge in spite of its comparatively high frequency and regularity. However, this structure is easy to learn as explicit knowledge because learners usually have no difficulty in understanding the grammatical rule: add -s to the base form of the verb in the 3rd person singular subject in the simple present tense. It is formally and functionally simple in terms of conceptual clarity and understanding and verbalizing the rule does not require much technical metalanguage. Long (2007) also emphasized the theoretical and pedagogical importance of investigating the influence that the salience and communicative value of a target structure have on learners’ ability to acquire it and suggested that recasts are “selectively facilitative” (p.107). In other words, he suggested that recasts and other implicit types of CF are more likely to work for the acquisition of meaning-bearing forms than for redundant structures.

Recent empirical studies examining the utility of CF for different target structures include Yang and Lyster (2010), Yilmaz (2012) and Li (2014). Yang and Lyster investigated the effects of corrective feedback on English past tense forms, reporting different effects for regular and irregular forms and concluding that these two forms are processed via distinct mechanisms. Yilmaz focused on one attribute of a target structure – saliency, positing that the Turkish plural is easier to hear and see (i.e. it has perceptual salience), that it has fewer allomorphs (i.e. demonstrates morpho-phonological regularity) and that it is more similar to the learners’ L1 than the Turkish locative. For these reasons he argued that the plural is easier to acquire as implicit knowledge than the locative. Li compared the Chinese perfective -le and classifiers, claiming that the former is perceptually less salient, has a neutral tone, is less transparent and functionally more complex than the latter. He concluded that it is more
difficult to acquire as both implicit knowledge and explicit knowledge. All of these studies provide evidence that the effectiveness of CF (in particular recasts) depends on the nature of target structure. These studies are examined in detail in Chapter 3.

However, there are to date too few studies that have examined the effect of CF in relation to the nature of the target feature to reach any conclusion on which attributes of target structure determine the efficacy of CF. Learning difficulty is not a clear-cut construct, especially when implicit and explicit knowledge are considered separately.

2.4 Implicit and explicit knowledge

Implicit knowledge is intuitive and procedural, easy to access and use but difficult to verbalize (R. Ellis, 2005b). It consists of memorized formulaic chunks and internalized constructions. Both categories of implicit knowledge are “largely hidden” and held unconsciously and only become apparent when learners are using the language (R. Ellis, 1994). This type of unanalyzed knowledge accounts for language fluency.

“Explicit L2 knowledge is that knowledge of rules and items that exist in an analyzed form so that learners are able to report what they know” (R. Ellis, 1994, p. 702). It includes “declarative knowledge” and “metalinguistic knowledge”, which can be “accessed through controlled processing when L2 learners experience some kind of difficulty in the use of the L2” (R. Ellis, 2004, p.245).

It is important to develop separate measures of the two types of knowledge. As R. Ellis (2005) argued, in order to investigate the process of acquisition it is essential to define and operationalize these two types of knowledge. For instance, explicit CF is found to facilitate language learning more effectively than implicit CF. A close examination shows, however, this finding is based more on the tests of explicit knowledge than the tests of implicit knowledge. Whether this conclusion holds true for implicit knowledge is not clear. Explicit CF enhances linguistic consciousness (Swain, 1995), makes the corrective function of CF clear, locates the error, and thus is more likely to trigger cognitive comparison (Ellis et. al., 2006). These researchers claimed that explicit CF directly or indirectly contributes to the development of implicit knowledge. In contrast, Long (1996) doubts this position, assuming that no form-meaning mapping can be achieved as the explicit techniques draw attention away from communication and treat language as a learning object. To confirm whether explicit CF contributes to language competence and whether it is more efficient than implicit
CF, it is necessary to design and implement measures of implicit knowledge so that potential testing bias can be overcome, as R. Ellis (2005; Ellis et. al, 2006) remarked. Meta-analysis studies (e.g. Li, 2010; Lyster & Saito, 2010; Mackey & Goo, 2007) have indicated that implicit CF involving recasts overall are more facilitative of implicit knowledge than of explicit knowledge. However, to what extent the benefits of implicit and explicit recasts vary depending on the kind of test used to measure learning is little researched.

In order to operationalize implicit and explicit knowledge, R. Ellis (2004; 2005b) identified seven distinct features: conscious awareness, type of knowledge, systematicity and certainty of L2 knowledge, accessibility of knowledge, use of L2 knowledge, self-report and learnability. For example, implicit knowledge refers to the linguistic forms that learners are intuitively aware of, which exist in the form of procedural knowledge and are variable but systematic, which can be accessed by means of automatic processing during fluent performance, which learners cannot verbalize and which are potentially acquired within the critical period. Measures of implicit knowledge require learners to respond by feel under time pressure so that the responses are of high consistency and certainty. So far, tests designed for implicit knowledge include elicited imitation, oral production, speeded dictation and timed grammaticality judgment tasks and those for explicit knowledge consist of untimed grammaticality judgment tasks, error correction tasks and metalinguistic tests (Révész, 2012).

2.5 Role of working memory in corrective feedback

Working memory is a mechanism for the temporary storage and manipulation of information that underpins complex cognitive tasks. It is different from short-term memory, which only retains information passively. WM models differ in terms of domain specificity and the number of their components, for instance, “Just and Carpenter’s domain-specific single-resource model, Baddeley’s and Waters and Caplan’s domain-specific multiple-resource model, [and] the domain-free connectionist models” (Sagarra, 2012).

The unitary construct of WM assumes that both processing and storage consume the same resources (e.g. Just & Carpenter, 1992). When a task demands more than the maximum of activated resources, an allocation scheme starts to work, resulting in the trade-off relationship between processing and storage. On the basis of such a model, working memory is usually tapped by means of complex span tests. For example, Daneman and Carpenter’s (1980) reading span test requires learners to read increasingly longer sets of sentences and recall the sentence-end words after finishing each set.
Baddeley’s (2000) WM is a non-unitary construct, which consists of the central executive, the phonological loop (for verbal and acoustic information), the visuo-spatial sketchpad (for visual and spatial information), and the multimodal episodic buffer (for integrating all sources of information into single units). The central executive is responsible for the overall control and allocation of attention by way of focusing, dividing and switching attention. This regulation and control function is measured by means of complex span tests, for example, reading span, listening span and speaking span tasks. The phonological loop is a slave system to serve the central executive and it temporarily stores acoustic information. It is described as having two components: phonological storage and articulatory rehearsal. Verbal information is translated into phonological form and reactivated via rehearsal before fading away so that the information can be maintained for a short period of time. It is this system that represents phonological short-term memory (PSTM). The PSTM is assessed by means of simple span tasks, in which learners are asked to listen to and recall sequences of unrelated items, such as non-words, words or digits. The central executive and the phonological loop are assumed to be involved in language learning and so are often addressed in the field of SLA.

The capacity of WM can be conceptualized in terms of the attentional resources that are available (Just & Carpenter, 1992). Attention functions as the “pivotal point” (Schmidt, 2010, p.735) for language acquisition, where learner external and internal factors meet. Whether learners have sufficient mental resources available in WM for the registration and processing of input determines the extent to which language is learned (Sawyer & Ranta, 2001). That is why researchers claimed that WM is implicated in L2 learning. This notion is evident in Robinson’s definition of “noticing” as a mechanism of “detection plus rehearsal in short-term memory” (1995, p.296). In his opinion, detection without awareness is not sufficient, and rehearsal in working memory leads to some level of awareness, which is necessary for acquisition. Conscious noticing constitutes a key way in which WM capacity affects language learning.

Working memory has been incorporated into the construct of language aptitude. Skehan (2002, p.91) posited that working memory is involved in the “noticing” and “pattern identification” stages of acquisition. Noticing of linguistic form depends not only on “phonemic coding ability” but also the management of attentional resources in WM. Whether learners successfully identify linguistic patterns in input is also related to WM as well as grammatical sensitivity. Miyake and Friedman (1998, p.339) claimed that “working memory
(WM) for language may be one (if not the) central component of this language aptitude”. They supported this relationship by arguing that the ability to keep separate but relevant pieces of information active simultaneously is an advantage for rule formation and meaning inferring.

Robinson (2005) and N. Ellis (2005) discussed the potential link between WM and recasts. Robinson stated that phonological short-term memory (PSTM) capacity and speed constrain whether a learner is able to retain the correct recast together with the learner’s erroneous utterance in WM long enough for subsequent cognitive comparison to take place, which in turn affects how much is learned from recasts. PSTM for maintaining the two utterances and analytic ability for noticing the gap jointly determine what learners gain from recasts. N. Ellis emphasized that complex working memory is crucial for recasts to work successfully. The advantage of recasts lies in presenting the correct form before the learner’s meaning is lost, so the span between learner utterance and the feedback forms a challenge to WM. Those learners who can maintain the meaning and simultaneously process the incoming information are more likely to benefit from recasts.

2.6 Contexts

Research has shown that the effectiveness of CF differs depending on whether the context is a classroom or a laboratory, and also between second and foreign language contexts. Different settings may motivate learners to attend more to either communicative meaning or linguistic form, and the orientation can have an effect on whether recasts promote acquisition. Sheen (2004) claimed that the differences in the rate of uptake across settings reflect learners’ orientation to form/meaning as well as the salience of the CF. In contexts where the learners are oriented to form and the recasts are explicit, for example, in the Korean EFL and New Zealand ESL classrooms in her study, repair is more likely to take place. Lyster and Mori’s Counterbalance Hypothesis predicts the variable effect of CF on acquisition. This hypothesis states that “instructional activities and interactional feedback that act as a counterbalance to the predominant communicative orientation of a given classroom setting will be more facilitative of interlanguage restructuring” (2006, p.294). According to this hypothesis, effective instruction varies depending on the context. In a form-focused setting, learners are “primed” (Lyster, 2007) to attend to the corrective function of recasts as they treat any input as an object to learn. On the other hand, in a meaning-focused classroom, recasts of the implicit type tend to be overlooked due to their homogeneity with the classroom activities.
It can be argued that classroom orientation is not a simple construct, especially in a task-based foreign language classroom. In task-based teaching, a setting is created that keeps learners primarily focused on meaning during the task performance. However, if the learners are not familiar with task-based teaching and naturally orient to form as a result of their previous classroom experiences, they may be more sensitive to CF. The extent to which such learners are oriented to form or meaning in task-based EFL lessons is uncertain.

### 2.7 Summary

Corrective feedback can be input-providing or output-pushing. It can also be explicit or implicit. Recasts constitute an input-providing type of CF but they can vary in explicitness. Corrective recasts are more explicit than implicit recasts and so more likely to provide learners both positive and negative evidence. Implicit recasts may only work as positive evidence as their corrective force is hidden and sometimes indistinguishable from non-corrective repetition.

SLA theories differ in whether they see CF contributing to language learning, which type of CF is more facilitative, and what mediates the effect of CF. Cognitive-interactionist theories contend that implicit CF (e.g. recasts) constitute what Long (1991) has called ‘focus on form’, which facilitates acquisition because it fosters joint attention to form and meaning. Recasts are assumed to trigger not only noticing of the form but also noticing-the-gap. In contrast, researchers drawing on the Output Hypothesis and the Skill-Learning Theory emphasize the importance of the negotiation of form through prompting learners to self-correct. Prompts are more likely to lead to uptake, which enhances control over the partially acquired knowledge. However, uptake of recasts is of low frequency. Corrective recasts are more likely to produce uptake than implicit recasts because their corrective force is clear to the learner but whether uptake of recasts is important to acquisition remains a controversial issue.

Goo and Mackey (2013) called for future research that focused on the ‘mechanism’ of recasts – that is how learners respond to them and how they affect the cognitive processes involved in L2 acquisition. One way of investigating this is by comparing recasts that vary in terms of how implicit/explicit they are. Drawing on Goo and Mackey’s proposal for future research, the study reported in this thesis will also consider attributes of the target structures investigated (e.g. their difficulty and salience), the effect that recasting has on both implicit knowledge and explicit knowledge by obtaining separate measures of each, the role played by
working memory (both phonological short-term memory and complex working memory and by contextual factors that influence whether learners are oriented to form or meaning.
Chapter 3  Empirical Studies on Recasts

Chapter 3 provides a brief review of empirical studies on the utility of recasts in terms of uptake, noticing and L2 learning. Some studies investigated the distribution of recasts in comparison to other types of corrective feedback (CF) and the rate of uptake in response to the CF in second language classrooms; some looked at the extent to which recasts trigger noticing using stimulated recall protocols or other introspective methods in laboratory settings; and others examined the effect of recasts on L2 learning as measured by different tests of linguistic knowledge. The relative effectiveness of recasts is reported separately for laboratory and classroom settings because it may be mediated by contexts as revealed in Li (2010) and Mackey and Goo (2007)’s meta-analyses. Meta-analyses of CF were conducted to see whether recasts work for L2 acquisition and what factors potentially constrain their acquisitional value. One of the moderators, working memory (WM), is focused on, and studies of the link between WM and recasts-driven L2 learning are reviewed in details. This chapter concludes with the findings from reviewing recent research on recasts.

3.1 Recasts and uptake

To judge whether CF contributes to language acquisition, we need to consider four criteria: existence, usefulness, usability and necessity (Pinker, 1989). 1. Do recasts occur in language teaching classrooms? 2. Are they useful for L2 learning? 3. Are they utilized in the ensuing utterances? 4. Are they necessary for L2 learning? This section focuses on classroom descriptive studies looking at the issues of “existence” and “usability” of recasts, and later sections will deal with the issue of “usefulness”. However, “necessity” of CF is an issue beyond the scope of the current study.

Lyster and Ranta (1997) developed a coding scheme for identifying six types of CF (explicit correction, recast, clarification request, metalinguistic feedback, elicitation and repetition) and two types of uptake (needs-repair and repair) to observe their distribution in Canadian French immersion classrooms. The researchers found that recasts comprise 55% of all the CF moves that the primary school participants received but only 31% of the recasts lead to uptake and 18% of the recasts produce repair. Panova and Lyster (2002) applied the taxonomy to their Canadian ESL classroom data, arriving at a similar pattern: recasts are most frequently provided (55%) to Canadian adult ESL learners, and 40% of the recasts result in uptake and 13% are followed by repair. On the basis of the low frequency of
successful uptake (i.e. repair), Lyster suggested that language teachers should consider using other types of feedback (i.e. prompts) more often rather than rely predominantly on recasts. For example, elicitation is most successful in terms of uptake in his studies.

However, Ellis et al. (2001) argued that language teachers do not need to be concerned about the intensity of recasts according to their observation of adult learners in intensive ESL communicative programs in a New Zealand private language school. They investigated the distribution of uptake in both pre-emptive and reactive focus-on-form episodes (FFE). About 82% of the responding FFEs contained recasts, a larger distribution than in previous studies. Their learners took up 72% of recasts and produced repair following 55% of recasts. These findings indicate that the learners in English language lessons tend to respond to recasts with a higher rate of repair due to their high concern for form.

Sheen (2004) examined the impact of contexts on the frequency of corrective feedback and uptake, drawing on the above mentioned studies and her own data. She collected the Korean EFL data from communicative English lessons in private language schools. It was found that the teachers recast errors more frequently in the Korean EFL and New Zealand ESL classrooms than in the Canadian contexts (ESL and immersion), and that the recasts invited a greater rate of uptake/repair in the former contexts. The adult Korean EFL learners produced the highest level of uptake (83%) and repair (58%) among the four communicative contexts. It seems that recasts are more successfully taken up in form-oriented communicative classrooms than in meaning-oriented classrooms.

Lyster and Mori (2006) compared the frequency of uptake following recasts between French and Japanese immersion settings. In both settings, recasts were the most widely used CF, but the response patterns were fundamentally different. A higher uptake and repair rate was found in response to prompts in the Canadian French immersion classrooms (Lyster & Ranta, 1997) whereas a higher level of uptake (61%) and repair (50%) followed recast moves in the American Japanese immersion classroom (Mori, 2002, as cited in Lyster & Mori, 2006), which was characterized by an analytic orientation. This comparison led them to propose the Counterbalance Hypothesis, which states that CF types that are most different from the classroom communicative orientation tend to be more effective for language learning than those that are congruent.
However, Yang (2009) observed that recasts are not the most frequently used type of CF (31%) in Chinese EFL classrooms, where prompts are predominant. She ascribed the prevalence of prompts over recasts to the English teachers’ heavy emphasis on formal accuracy. They dealt with linguistic forms independently from communicative activities. The instruction relied more on negotiation of form than on recasts and resulted in a high rate of uptake overall, but seldom left an opportunity for uptake following recasts (27%). However, when there was an opportunity for uptake following a recast, the learners repaired their errors.

<table>
<thead>
<tr>
<th>Study</th>
<th>Context</th>
<th>Participants</th>
<th>Recast rate (% of all CF)</th>
<th>Uptake rate (% of recasts)</th>
<th>Repair rate (% of recasts)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lyster &amp; Ranta (1997)</td>
<td>Canadian French immersion</td>
<td>Primary school; intermediate</td>
<td>55%</td>
<td>31%</td>
<td>18%</td>
</tr>
<tr>
<td>Ellis et al. (2001)</td>
<td>New Zealand ESL</td>
<td>Adult; (low) intermediate</td>
<td>82%</td>
<td>72%</td>
<td>55%</td>
</tr>
<tr>
<td>Panova &amp; Lyster (2002)</td>
<td>Canadian ESL</td>
<td>Adults; beginners</td>
<td>55%</td>
<td>40%</td>
<td>13%</td>
</tr>
<tr>
<td>Sheen (2004)</td>
<td>Korean EFL</td>
<td>Adult; (high) intermediate</td>
<td>83%</td>
<td>83%</td>
<td>58%</td>
</tr>
<tr>
<td>Mori (2002)</td>
<td>American Japanese immersion</td>
<td>Primary school</td>
<td>65%</td>
<td>61%</td>
<td>50%</td>
</tr>
<tr>
<td>Yang (2009)</td>
<td>Chinese EFL</td>
<td>Secondary school</td>
<td>31%</td>
<td>27%</td>
<td>27%</td>
</tr>
</tbody>
</table>

What constrains uptake rate after recasts has been addressed in terms of the nature of the CF, the target structure, opportunities for response, learner factors and contextual variables. For example, Lyster (1998b) claimed that recasts are ambiguous between form correction and meaning confirmation in Canadian immersion classrooms. Failure to recognize the corrective intent explains in part the low level of repair.

Sheen (2004) posited that the higher uptake rate in Korean EFL and NZ ESL than in the Canadian contexts derived from the learners’ form-focused orientation and the salience of the
CF, which was typically very brief. Then Sheen (2006) concentrated on the link between the characteristics of CF and immediate uptake, concluding that length, linguistic focus, type of change, mode, use of reduction and number of changes affect the rate of repair. She labelled short, partial, repeated, declarative recasts, targeting a single structure and correcting via substitution, as “explicit recasts”. Her finding corroborates Nicholas, Lightbown and Spada’s position that “recasts appear to be most effective in contexts where it is clear to the learner that the recast is a reaction to the accuracy of the form, not the content, of the original utterance” (2001, p. 720).

The Counterbalance Hypothesis underscores the significance of contextual variables in determining learners’ perception of CF. The same type of CF in various instructional settings is likely to lead to different levels of uptake (Ellis & Sheen, 2006).

As for target structures, Lyster (1998a) reported that recasts are more likely to generate successful uptake when they target phonological than lexical or grammatical errors in immersion classrooms. In contrast, no significant difference was found between the target features in Ellis et al.’s (2001) study and linguistic features did not predict the success of uptake in Loewen’s (2004) study of FFEs either. Instead, these researchers attributed the higher level of uptake to the learners’ orientation to form in addition to the characteristics of FFEs (reactive, complex and involving negotiation of form) because their participants were fee-paying international students enrolled in language classes (ESL).

Oliver and Grote (2010) attributed the low uptake rate to another learner factor – age. They asserted that the child ESL learners in their study were cognitively immature, which prevented them from focusing on form and attending to the recasts. They had difficulty shifting from meaning to form. However, this claim is not supported by the results shown in Table 2. For instance, adult learners in Panova and Lyster (2002) produced fewer repairs whereas primary school students in Mori (2002) had a high level of repair.

Low uptake rate following recasts can also result from lack of opportunities for output, as Oliver (2000) and Yang (2009) noted. The learners had no chance to use the CF one third of the time in the teacher-fronted class that Oliver observed. The teacher continued talking without expecting any response from learners. After excluding the instances of topic continuation, Yang found that her learners’ uptake of the recasts was all successful due to the
explicitness of the CF. The pedagogical significance of this finding is that teachers need to leave time for students to uptake when implementing recasts if uptake of recasts matters.

Now I will return to Pinker’s terms, “existence” and “usability”, to see what the classroom descriptive studies reviewed in this section tell us about the existence and usability of recasts. With regard to existence, they indicate overall that recasts constitute the most frequently used CF strategy in meaning-oriented and form-oriented communicative instruction. Although prompts were dominant in the form-focused instruction in Yang (2009), the frequency of recasts in this study were still comparatively high (31%). Recasts figure frequently in second language classrooms. With regard to usability, uptake and repair rates in response to recasts vary according to context: they are lower in meaning-oriented classrooms but higher in form-oriented classrooms when learners have an opportunity for output following the CF. Short and partial recasts targeting a single structure are more explicit and lead to more successful uptake. However, there is limited evidence to show that target structure or learners’ age affects the extent to which the learners uptake and repair their errors.

Repair in response to CF is often taken as a piece of evidence for noticing, however, not all noticing is captured by repair. It is possible that a learner has noticed a recast but had no opportunity or chose not to uptake because the recast provided the correct form. As uptake is not a reliable measure of noticing, online and retrospective recall protocols have been widely used in later studies to investigate noticing.

3.2 Recasts and noticing

There are three levels of noticing: noticing the form in the input, noticing the corrective intent and noticing the gap between the learner’s interlanguage and the target language (Sakai, 2011). Noticing of recasts in this section refers to perception of the corrective function of the CF (negative evidence) or perception of the locus of errors (noticing the gap). Noticing of CF has been measured in interaction studies by means of retrospective recall (e.g. Mackey, Gass, & McDonough, 2000; Mackey, Philp, Egi, Fujii, & Tatsumi, 2002), online recall (e.g. Kartchava & Ammar, 2014; Philp, 2003; Trofimovich, Ammar, & Gatbonton, 2007) and eye tracking (Smith, 2012) in addition to uptake. Some studies used a combination of these measures to tap noticing. For instance, Mackey (2006) employed online learning journals, stimulated recall and questionnaire responses, finding that CF overall contributed to noticing. In her classroom setting, recasts were frequently provided for plurals and past tense errors,
but the learners reported fewer instances of noticing of these structures. This can be explained by the non-salience of the target structures along with the implicitness of the CF.

Noticeability of recasts is rarely compared to that of other forms of CF targeting the same linguistic feature except in Kartchava and Ammar (2014). They conducted a classroom-based study of recasts, prompts and a combination of the two in leading to noticing (and L2 learning). Noticing was measured by means of immediate recall. The learners reported less noticing of recasts than of prompts and of the combined CF on past tense errors. Prompts pushed the learners to produce responses and this pushing may have rendered the corrective function more salient. The combined CF is quite similar to Doughty and Varela’s (1998) “corrective recast”, which is comprised of a repetition of the error and a recast. The prompt move diverts the learner’s attention from meaning to form, and makes him or her ready for the ensuing recast move. The findings indicate that complex recasts trigger more noticing than simple ones although no significant difference has been found in L2 learning. Kartchava and Ammar (2014) also found that noticeability of CF varies depending on the target structures, with more noticing of past tense than of question formation. The high frequency of irregular past forms (74%), which are salient, may account for the difference in noticing between the two linguistic features.

Previous studies of noticeability of CF have generally focused on recasts alone. Recasts are assumed to be implicit in nature and tend to be overlooked or misperceived as confirmation checks. To what extent recasts are perceived as corrective feedback has been investigated and the variables mediating the noticeability of recasts have been identified as length of recasts (Egi, 2007; Philp, 2003), type of errors (Gass & Lewis, 2007; Mackey et al., 2000), learner proficiency (Philp, 2003; Trofimovich et al., 2007) and working memory (Trofimovich et al., 2007; Mackey et al., 2002).

Philp (2003) drew on cued immediate recall to investigate the extent to which 33 adult ESL learners noticed recasts on question forms in NS-NNS dyadic interaction. It was found that 60% to 70% of the recasts were recalled accurately. High and intermediate level learners noticed recasts more than low level learners because the lack of linguistic knowledge and unfamiliarity with the target structure may have constrained the latter’s interpretation of the CF. Shorter recasts with fewer changes can be well maintained in WM, available for deeper processing, and thus are more likely to be retrieved. This is in line with Egi’s (2007) findings. Her immediate reports and stimulated recall protocols showed that 49 learners of Japanese as
a foreign language interpreted short recasts resembling the original utterances often as linguistic corrections rather than as responses to content in NS-NNS communicative tasks. There was no significant difference between morphosyntactic and lexical recasts in terms of reported noticing (61% and 57%) and interpretation (79% and 82%) of the CF. The similarity of pattern was attributed to the consistent focus of morphosyntactic recasts, which targeted mainly the morpheme -te and numeral classifiers, and the various targets of lexical recasts.

In stark contrast, Mackey et al. (2000) reported that the ESL and Japanese as a foreign language learners had more accurate perception of lexical and phonological than morphosyntactic CF based on their stimulated recall comments on the CF episodes in task-based dyadic interaction. A closer examination showed that morphosyntactic errors were often corrected via recasts (75% of recasts for morphosyntactic errors). The inaccurate perception of morphosyntactic CF was attributed to the implicitness of recasts as well as the low communicative value of morphosyntactic forms. Although the morphosyntactic CF was misperceived as lexical or phonological CF, its corrective intent was consistently recognized by learners of Japanese as a foreign language (68%). Gass and Lewis (2007) replicated Mackey et al.’s study distinguishing heritage language and non-heritage language learners and reached a similar conclusion that morphosyntactic CF was less accurately perceived than phonological and lexical CF by both types of learners. Recasts were provided to non-heritage language learners most often for their morphosyntactic errors. These learners did not accurately judge the type of errors (35% in accuracy) but noticed the corrective force of morphosyntactic CF 73% of the time.

Other researchers focused on the moderating role of learner cognitive factors on noticeability. For instance, Mackey et al. (2002) adopted stimulated recall and exit questionnaire methods and found that reported noticing of recasts was positively related to the learners’ working memory (WM) capacity as illustrated by composite scores, but not to subtest scores for phonological short-term memory (PSTM) or complex working memory (CWM). The capacity of WM constrains noticing of CF during dyadic interaction, but noticing is not determined by WM alone. The high PSTM learners’ noticing also varies depending on proficiency, with learners at a low developmental level showing more noticing. The high proficiency learners (Stage 5) responded less to recasts because they had acquired the target structure. Trofimovich et al. (2007) examined the role of phonological memory, working memory, attention control and analytical ability in the noticing of recasts provided
individually via computers, but found that none of these four cognitive factors were associated with noticing. They speculated that the recasts provided via computers were salient and predictable, which the learners attended to without relying too much on cognitive resources. It is possible that learner factors are involved in the noticing of recasts to a different extent depending on whether the CF is more implicit or explicit. Their study also confirmed the mediation of linguistic features: grammatical targets were less frequently detected than lexical targets.

Taken together, these studies suggest that recasts in classroom settings are not easy to notice but recasts clearly enhance noticing under laboratory conditions. Recasts in combination with other forms of CF are more explicit than recasts alone, leading to higher recognition of corrective intent (e.g. Kartchava & Ammar, 2014). Short recasts involving few changes are more likely to be noticed (as shown in Philp, 2003; Egi, 2007). The noticeability of explicit recasts is reported irrespective of setting, but is unlikely to be affected by learner factors (e.g. Trofimovich et al., 2007). The noticeability of implicit recasts depends on the salience of target feature (Kartchava & Ammar, 2014), learner proficiency (Philp, 2003; Trofimovich et al., 2007), the learners’ form-orientation (JFL learners in Mackey et al., 2000; non-heritage language learners in Gass & Lewis, 2007) and working memory capacity (Mackey et al., 2002). These various factors mediate the noticeability of implicit recasts. To examine how the nature of CF, salience of target feature and learner variables affect the noticeability of recasts, more empirical studies are needed. Future research on the noticeability of recasts needs to adopt a variety of tests to collect more information about noticing, as in Mackey (2006), considering the fact that not all noticing is reported in online and retrospective recall protocols.

The research indicates that learners are more likely to recognize the corrective intent of recasts in laboratory settings, but there is uncertainty as to whether the nature of the linguistic feature reformulated in recasts (i.e. whether phonological, lexical or morphosyntactic) affects whether or not they are attended to (Egi, 2007, cf Mackey et al., 2000). In other words, it is unclear whether the noticing of the focus of recasts varies according to the type of error.

3.3 Recasts and acquisition

Linguistic knowledge tests have been used to measure the contribution of recasts to second language acquisition. There were two ways to address the relationship between recasts and acquisition. Some (quasi-)experimental studies simply looked at the effect of recasts on L2
learning as measured in posttests while other empirical studies examined the effect of recasts on uptake and the relationship between uptake and acquisition.

3.3.1 Recasts and acquisition in terms of posttest scores

Empirical studies on the usefulness of recasts for language learning have been conducted to investigate whether recasts drive language learning, whether the CF does so via negative evidence, and whether explicitness makes a difference to the effect of the CF.

Doughty and Varela (1998) operationalized explicit recasts as two CF moves: a repetition of the error and a recast, attempting to integrate focus on form into science instruction for middle school ESL learners. The recast group had significant and durable gains in the use and accuracy of past and conditional forms in the oral and written tests (except for the gain in accuracy in the written delayed posttest). However, the comparison group did not change much except for a gain in the written posttest. Explicit recasts in their classroom setting study contributed to the learning of past time reference.

Mackey and Philp (1998) conducted an experimental study on the effect of recasts distinguishing ready from unready ESL learners. They found that recasts worked better for the developmentally ready learners and that recasts worked better than interaction without CF in promoting the production of advanced question forms. If a target structure was “beyond the learner’s current stage of development” (Ellis & Sheen, 2006, p. 591), he or she could not benefit from the recasting even if the feedback was intensive.

Han (2002) recruited eight upper-intermediate ESL learners, who were assumed to be developmentally ready for acquiring tense consistency, to investigate the effect of recasts in a laboratory setting. The recast group gained better control of tense consistency than the comparison group as illustrated by their oral and written narratives. Han argued for the utility of recasts in enhancing awareness and producing tense consistency but cautioned that this benefit needed to be tested on other target structures.

With a larger sample (74 learners of L2 Spanish), Leeman (2003) confirmed the facilitative role of recasts in learning L2 grammar. She claimed that recasts not only promote the development of structures of communicative value, but also of non-salient redundant forms, for example, noun-adjective agreement in gender and number. The learners receiving recasts and those receiving enhanced positive evidence significantly outperformed the control group but the learners receiving negative evidence did not. She asserted that negative evidence may
not be the crucial element of recasts. Put another way, recasts are beneficial due to the positive evidence they provide even if the negative evidence fails to work.

Explicit recasts are more likely to be perceived as negative evidence than implicit recasts. The comparison between the two types of recasts may shed light on how recasts facilitate language learning. Erlam and Loewen (2010) investigated the effect of explicit and implicit recasts on French as a foreign language university learners’ accuracy in noun-adjective agreement in spontaneous production, elicited imitation and untimed grammaticality judgment tests (GJT). The explicit and implicit recasts had a similar impact on the three measures, and neither of them revealed an advantage over the communicative tasks alone. The implicit recasts in their laboratory setting were salient and were noticed (67%) to a similar extent as explicit recasts (62%). In fact, the difference in CF was further overshadowed by the considerable effect of task-based interaction alone. About 44% of the comparison group were aware of the target structure and they achieved greater gains in the use of the target structure.

However, the explicit recast group in Chen (2010) outperformed the implicit recast and the comparison groups, although all the groups improved their accuracy in noun plural over time. The superiority of explicit recasts in the GJT and metalinguistic tests seems to indicate that this type of CF affords negative evidence, which is facilitative of language learning, at least, in terms of explicit knowledge. The overall gains over time, including the comparison group, may again be attributed to the Chinese EFL laboratory setting, where the learners’ tendency to a form-orientation enhances the benefits of both interaction and recasts.

Now I would like to synthesize the above-mentioned studies to answer the questions raised at the introduction to this section. As regards the usefulness of recasts, explicit recasts can drive L2 grammar learning irrespective of settings. The only exception is Erlam and Loewen’s (2010) study, in which the treatment lasted only for one hour. As the overall improvement from pretest to posttest demonstrates, it is possible that a longer treatment would have differentiated the effectiveness of explicit recasts from the comparison condition. A question remains about the efficacy of implicit recasts although they are rendered more salient in lab settings. As to whether recasts contribute to language learning through negative evidence, there is no consensus. Leeman (2003) doubted the value of recasts as negative evidence whereas Chen (2010) argued for the beneficial role of negative evidence that explicit recasts provided. Doughty (2003) also contended that “the implicit negative evidence provided to
learners by recasts, contingent upon their interlanguage utterances, is noticed and used in SLA” (pp. 289-290). As regards the explicitness in recasts, there are too few studies comparing explicit and implicit recasts to reach a firm conclusion. Erlam and Loewen (2010) found no evidence for any advantage of explicit over implicit recasts while Chen (2010) reported that explicitness in recasts played a role in the development of explicit knowledge. Nassaji (2009) also claimed that explicitness makes a greater difference to the effect of recasts than prompts. Explicit recasts facilitated more short-term and long-term learning than implicit recasts in his study, which focused on the comparison between recasts and elicitations and so will be reviewed in Section 3.4.2. To resolve the uncertainty about the role of explicitness in recasts, future comparative studies need to utilize different tests for different target structures in classroom settings, given the impact of methodological and contextual factors.
<table>
<thead>
<tr>
<th>Study</th>
<th>Participants</th>
<th>Target structure</th>
<th>Instruction</th>
<th>Tests</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Doughty &amp; Varela (1998)</td>
<td>Intermediate ESL students studying science</td>
<td>Past and conditional forms</td>
<td>Form-focused instruction</td>
<td>Oral and written experimental reports</td>
<td>recasts &gt; comparison in gains</td>
</tr>
<tr>
<td>Mackey &amp; Philp (1998)</td>
<td>Beginner, lower-intermediate adult ESL private school learners</td>
<td>Question forms</td>
<td>Story completion, picture sequencing and picture drawing</td>
<td>Spot-the-difference tasks</td>
<td>Recasts &gt; comparison for developmentally ready learners</td>
</tr>
<tr>
<td>Han (2002)</td>
<td>Adult ESL upper intermediate learners</td>
<td>Tense consistency</td>
<td>Story narration</td>
<td>Oral and written narratives</td>
<td>Recasts &gt; comparison</td>
</tr>
<tr>
<td>Leeman (2003)</td>
<td>1st-year university learners of Spanish</td>
<td>Noun-adjective agreement</td>
<td>One-on-one tasks</td>
<td>Picture description tasks</td>
<td>Recasts = enhanced salience &gt; control</td>
</tr>
<tr>
<td>Erlam &amp; Loewen (2010)</td>
<td>2nd and 3rd-year university learners of French</td>
<td>Noun-adjective agreement</td>
<td>communicative tasks with a group of learners</td>
<td>Spontaneous production, elicited imitation, and untimed GJT tests</td>
<td>Explicit recasts = implicit recasts = comparison</td>
</tr>
<tr>
<td>Chen (2010)</td>
<td>Primary school learners of EFL</td>
<td>Noun plural</td>
<td>Information gap tasks</td>
<td>GJT and metalinguistic knowledge tests</td>
<td>Explicit recasts&gt;implicit recasts=comparison</td>
</tr>
</tbody>
</table>
3.3.2 Recasts, uptake and acquisition

Uptake rate has been used as one way of evaluating the contribution that recasts make to acquisition. Lyster and colleagues (Lyster & Ranta, 1997; Panova & Lyster, 2002) argued that recasts are less effective than prompts based on uptake rate following the two types of CF. They acknowledged that uptake does not equate to acquisition, but claimed that uptake contributes to L2 learning. However, Mackey and Philp (1998) argued against the uptake-acquisition link, claiming that uptake of recasts is a “red herring”. Whether repetitions of recasts are of value to L2 learning has been an object of enquiry for a number of empirical studies.

Loewen and Philp (2006) analyzed how the characteristics of recasts affected successful uptake and posttest performance in adult ESL classrooms. Predictive of uptake were stress, declarative intonation, one change and multiple feedback moves, which rendered recasts more explicit. Indicators of posttest scores were interrogative intonation, shortened length and one change. In their study, uptake of elicitation moves predicted L2 learning but repetition following recasts did not.

In laboratory settings, the findings concerning the uptake-acquisition relationship are mixed. McDonough and Mackey (2006) examined how Thai university learners of English as a foreign language responded to recasts in dyadic interaction and to what extent they manifested development of question forms in posttests. The learners’ primed production of the target structure (i.e. a new utterance using the target structure across subsequent moves) was a predictor of L2 learning but immediate repetition of recasts was not. McDonough (2007) also argued for the view that uptake of recasts and learning outcomes were not associated. In her comparative study on the effect of recasts and clarification requests (see Section 3.4.2 for details), the English as a foreign language learners took up only 15% of the recasts when provided with opportunities to do so, but they benefited from recasts as much as clarification requests as measured by development of past tense verbs in the posttest. In Révész, Sachs, and Mackey’s (2011) study, however, the EFL learners’ uptake of recasts was found to predict their development of past progressive forms in a condition where there was visual support (i.e. seeing the picture while describing) but it was not predictive in a condition without visual support. They claimed that task complexity moderates the link between uptake and learning outcomes. It should be noted that their recasts were short and salient and their tasks were tightly controlled. Such a design may have rendered recasts
effective at both the level of uptake and of learning as measured by the posttests, and thus have resulted in a positive uptake-learning relationship.

In summary, the above studies indicate that immediate repetition following recasts is not related to L2 learning in classroom settings, where different variables may constrain learners’ ability to benefit from recasts. The uptake-acquisition relationship is sometimes evident in laboratory settings. This can be explained by a variety of factors. 1). Primed production is a better predictor of learning outcome than immediate repetition following recasts. 2). Successful uptake may be predictive when recasts are explicit because the explicit recasts trigger both immediate uptake and subsequent learning. 3). Tasks that are more demanding of attentional resources may influence whether there is a relationship between uptake and acquisition following recasts.

**3.4 Recasting vs modelling vs prompting and acquisition in laboratory settings**

The acquisitional value of recasts has also been examined in comparison to other forms of CF. Some studies looked at the relative effect of recasting as opposed to modelling, and others considered the relative effect of recasting as opposed to prompting. Some comparative studies were conducted in laboratory settings while others occurred in classroom settings. I am going to begin with the laboratory-based comparative studies in this section. The classroom-based comparative studies will be reviewed in next section.

**3.4.1 Recasting vs modelling in laboratory settings**

Recasts are assumed to consist of different features – positive evidence, negative evidence and enhanced salience (Leeman, 2003) – whereas models provide positive evidence only. The comparison between the two types of input allows researchers to observe the role of negative evidence and saliency in recasts.

Long et al. (1998) conducted two experimental studies of the relative effect of recasts and models, one on learning Japanese adjective ordering and locatives and the other on acquiring Spanish object topicalization and adverb placement. The young adult participants did not show a clearly consistent pattern for the Japanese structures due to their prior knowledge and the opportunities created for repeating the models, which led to attention to form. Recasts were found to be more facilitative than models for Spanish adverb placement, but not for Spanish object topicalization. The last structure was beyond the learners’ ability to learn. The
results indicate that recasts promote grammar learning but their benefit differs according to the target structure.

Iwashita (2003) compared the role of positive (models) and negative (mainly recasts: 70%) evidence in the immediate development of Japanese grammatical structures in a quasi-experiment. Models were observed to be ten times more frequent than negative evidence in the NS-NNS interaction, but benefited only high level learners. Negative CF including recasts was found to be facilitative irrespective of learner proficiency. Recasts were more predictive of the short-term gains of the te-form verbs than other conversational moves but they did not have impact on the development of locative-initial construction. The efficacy of recasts varied according to the target structure. She argued that the salience and intensity of CF contribute to L2 learning as well as frequency.

As far as the few studies that have compared recasts and models, I will tentatively address the question of the relative effect of recasts. Recasts seem to work better than models for L2 grammar learning in laboratory-based settings. The comparatively greater impact of recasts depends on the target structures. Recasts are effective for those linguistic features for which learners are developmentally ready but not for those for which learners are unready. When learners are developmentally ready, recasts benefit both high and low proficiency learners (as indicated by their prior knowledge) in laboratory settings.

3.4.2 Recasting vs prompting in laboratory settings

Recasts provide a correct form of the erroneous utterance whereas prompts withhold the correct form but elicit self-repair from learners. The comparison between recasting and prompting helps us to gain a better understanding of the relative importance of input-providing and output-prompting strategies in language acquisition.

McDonough (2007) controlled for the effect of the salience of the CF by comparing recasts and clarification requests, both of which were implicit in nature. She carefully selected 74 EFL learners to ensure learner readiness for learning simple past activity verbs. After the treatment, more past tense verbs were learned in the recast and clarification request groups than in the comparison group, pointing to the efficacy of CF moves. There was no significant difference between the two types of CF. Clarification requests, as an implicit form of prompts, may not necessarily be superior to recasts.
Loewen and Nabei (2007) examined recasts against two types of prompts – clarification requests (implicit) and metalinguistic CF (explicit) to differentiate the contribution of CF explicitness and self-repair opportunities. They found a positive effect for all the CF types in the timed grammaticality judgment test (GJT), but not in the untimed GJT and oral production tests. The three types of CF may have facilitated the EFL learners’ implicit knowledge of question formation only in detecting grammaticality but not in producing the target structure due to the short length of treatment (half an hour). No difference between CF types was found in their study. It is possible that the CF given to small groups of learners, including recasts, was relatively individualized and thus became more noticeable. That is to say, recasts in laboratory-like classrooms are salient.

Nassaji (2009) compared the benefits of recasts and another subcategory of prompts – elicitations – using post-interaction error identification and correction tasks. He found that ESL learners receiving recasts performed better than those receiving elicitations in the correction tests. The more explicit form in either case (recasts or elicitations) led to more immediate and delayed learning than the implicit counterpart and the effect of explicitness was greater for recasts. These findings suggest that explicitness is a key factor in recasts. His recasts were given incidentally, focusing on various linguistic forms, some of which were unknown to the learners. It can be speculated that CF works in different ways. Recasts benefit the learning of new linguistic features whereas elicitations promote the proceduralization of existing declarative knowledge.

Lyster and Izquierdo (2009) held the same position that CF works differently. Their learners of French as a foreign language received 3-hours of form-focused instruction prior to the CF treatment in dyadic interaction with (near) native speakers of French. The recast receivers improved their accuracy of French grammatical gender because the CF repeatedly provided correct exemplars together with opportunities to infer what was incorrect, whereas the prompt receivers benefited from the negative evidence and opportunities for modified output that the CF afforded. Both recasts and prompts promoted the development of the target structure and there was no significant difference between the two types of CF. Arguably, the form-focused instruction prior to the treatment primed the learners to attend to recasts as well as to prompts.
<table>
<thead>
<tr>
<th>Study</th>
<th>Participants</th>
<th>Target structure</th>
<th>Instruction</th>
<th>Tests</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>McDonough (2007)</td>
<td>1st-year university EFL learners</td>
<td>Simple past activity verbs</td>
<td>Information exchange and information gap tasks</td>
<td>Oral production</td>
<td>Recasts&gt; comparison; Recasts = clarification requests</td>
</tr>
<tr>
<td>Loewen &amp; Nabei (2007)</td>
<td>University EFL learners</td>
<td>English questions</td>
<td>1 spot-the-difference task and 1 guess-the-storyline task</td>
<td>Timed and untimed GJT and oral production tests</td>
<td>Recasts= metalinguistic = elicitation &gt; control and comparison on timed GJT</td>
</tr>
<tr>
<td>Lyster &amp; Izquierdo (2009)</td>
<td>University learners of French as a foreign language</td>
<td>French grammatical gender</td>
<td>Form-focused instruction + Object identification, picture description, and riddles</td>
<td>Oral production, binary choice and reaction time tests</td>
<td>Recasts = prompts</td>
</tr>
<tr>
<td>Nassaji (2009)</td>
<td>Adult ESL learners</td>
<td>Multiple forms</td>
<td>1 picture sequencing task</td>
<td>Identification and correction of erroneous utterances</td>
<td>Recasts &gt; elicitations; Explicit recasts &gt; implicit recasts</td>
</tr>
</tbody>
</table>
I have examined whether recasts facilitate L2 grammar learning in comparison to prompts, why they are effective in laboratory settings, and what implication the relative effects have for understanding the mechanism of CF. Together, these laboratory-based comparative studies generated a positive effect for recasts, which were as effective as or even more effective than (one form of) prompts. The laboratory setting, where variables are tightly controlled, meets the conditions for recasts to be effective. In laboratory contexts, as Han (2002) argued, learners are few, receiving “individualized attention” (p. 568); they are developmentally ready for the target structures; recasts consistently target one or few specific structures, raising learners’ awareness of correction; and the CF is delivered intensively, making it salient. Frequency, intensity and saliency are crucial moderators of input-providing CF. The similar effect of recasts to prompts in laboratories suggests that both types of CF work for L2 learning. Recasts provide both positive evidence and opportunities for inferring negative evidence and so may be more beneficial for learning new linguistic knowledge. By contrast, prompts provide negative evidence and opportunities for pushed output and thus are more facilitative of gaining control of existing knowledge.

3.5 Recasting vs prompting and acquisition in classroom settings

Classroom-based empirical studies only compared the efficacy of recasts against prompts but not against models. This section looks at the comparative studies to investigate whether recasts promote the development of L2 grammar as effectively as prompts in classrooms, and what variables mediate the effect of recasts.

Lyster (2004) conducted a classroom quasi-experimental study, in which he implemented recasts, prompts or no CF with French immersion fifth graders in addition to form-focused instruction. The FFI-prompt group significantly outperformed the FFI-recast, FFI-only and control groups in written tests for French grammatical gender. The FFI-recast group performed better than the control but not the FFI-only group. He interpreted the relative ineffectiveness of recasts in terms of form retrieval, feedback ambiguity, noticing difficulty, conscious awareness and output opportunity. 1) Learners simply heard the target forms in the recasts and did not have to retrieve them from long-term memory. 2) Sometimes, they could not distinguish the corrective recasts from the confirmative recasts. 3) Recasts of morphosyntactic errors were difficult to notice. 4) Prompts created “increased opportunities for conscious awareness of their teacher’s feedback” (p. 427). 5) Prompts pushed the learners
to produce while recasts were only receptive. However, neither FFI-prompt nor FFI-recast groups outperformed the FFI-only group in oral tests.

Ammar also used written and oral tests but found an overall superiority for prompts over recasts on oral tasks. Ammar and Spada (2006) discussed the relative advantage in terms of learners’ proficiency. For learners with high prior knowledge of 3rd person possessive determiners, prompts and recasts were equally beneficial, but for learners with little prior knowledge prompts were more effective. Their prompts were implemented stepwise: indicating an error and then providing metalinguistic information. The multiple-move prompts were more explicit and more likely to assist the low-proficiency learners, who had difficulty in identifying the presence and locus of an error, than recasts. The second advantage of prompts is that they create an opportunity for learners to produce self-repair. In Ammar’s (2008) study, prompts were again more effective than recasts in leading to L2 morphosyntactic development in oral tests. The advantage of prompts was especially evident for low-proficiency learners. The decreased reaction time in completing tasks demonstrates that prompts work better for the proceduralization of prior knowledge. These findings concur with Mackey and Philp’s (1998) conclusion that the efficacy of recasts is constrained by learner proficiency. Proficiency, among other factors, moderates the comparative efficacy of CF. Therefore, there is no single ideal CF that fits all learners (Ammar & Spada, 2006).

Yang and Lyster (2010) found that the comparative effects of prompts and recasts also depended on target structures. Prompts worked better than recasts for the development of regular past forms but the two types of CF had a similar effect on the acquisition of irregular forms. The rule-based regular forms may have biased prompts because the CF is salient and creates opportunities for retrieval of exemplars, and the repeated retrieval assists the internalization of the rule. The high salience, high frequency but low regularity of irregular past forms may have rendered the corrective function of recasts more noticeable and thus more beneficial than in the case of regular forms.

Sheen (2007) targeted English definite and indefinite articles in her comparison of recasts against a combination of explicit correction and metalinguistic explanation. She found that adult EFL learners who received correction-metalinguistic CF performed better than those who received recasts or no CF in both immediate and delayed posttests whereas the recast receivers did not outperform the control group. She suggested that the combination of explicit
correction and metalinguistic explanation led to a deeper level of processing, involving understanding as well as noticing.

Ellis et al. (2006) compared the effect of recasts as opposed to a combination of repetition and metalinguistic CF on the acquisition of past tense -ed. A general superiority of repetition-metalinguistic CF over recasts was reported in the delayed imitation and grammaticality judgment posttests. The recast group outperformed the control group only in the oral imitation delayed posttest. The impact of CF was more evident in the tests of implicit knowledge than explicit knowledge. The metalinguistic CF, comprising repetition and metalinguistic explanation, was more salient and so easier to be attended to than implicit recasts. It is likely to lead to greater depth of noticing (i.e. noticing the gap) and consequently more gains in implicit knowledge than recasts. R. Ellis (2007) examined the effect of metalinguistic CF and recasts on the acquisition of two different structures: past tense -ed and comparative -er. For both target structures, the recast group did not differ from the control group whereas the metalinguistic CF group did, pointing to an advantage for the latter CF. Again, his metalinguistic CF consisted of a repetition of the error and metalinguistic explanation for past tense -ed. This type of combined CF tends to be more beneficial for language learning. Metalinguistic CF’s immediate effect on comparative -er but delayed effect on past tense -ed indicate that its effectiveness depends on the target structure.

The above studies compared one particular type of CF (recasts) and a group of CF types (prompts). Recasts were manipulated as one single move but prompts sometimes as multiple moves. Such comparisons are weighted in favour of prompts (Goo & Mackey, 2013). Ideally, recasts need to be compared with just one type of prompts.

Operationalizing metalinguistic CF and recasts as single moves and continuing the topic without leaving an opportunity for uptake, Goo (2012) found that metalinguistic CF was not superior to recasts in promoting the development of English that-trace filter. The metalinguistic CF and recast groups both outperformed the control group but there was no significant difference between the two experimental groups. The results indicate that recasts are as effective as metalinguistic CF when there is no uptake opportunity. The unavailability of uptake had great impact on the efficacy of metalinguistic CF as its acquisitional value lies in pushing learners to modify their non-targetlike utterance. Working memory (WM) was found to moderate the effectiveness of recasts but not metalinguistic CF, suggesting that
recasts contribute to language learning when learners make a cognitive comparison in their WM.

Mifka-Profozic (2013) compared the effect of recasts and clarification requests (a single type of prompts) and implemented them as single moves on secondary school French-as-a-foreign-language learners’ grammar acquisition. She found that recasts were more facilitative than clarification requests for the short-term learning of passé composé and the long-term learning of imparfait in oral tests. Recasts also had an advantage over clarification requests in the long-term effect on the learning of both structures in written tests. The superiority of recasts was more evident for high-proficiency learners than for low-proficiency learners.
<table>
<thead>
<tr>
<th>Study</th>
<th>Participants</th>
<th>Target structure</th>
<th>Instruction</th>
<th>Tests</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lyster (2004)</td>
<td>Fifth graders in French Immersion classes</td>
<td>French grammatical gender</td>
<td>Form-focused instruction</td>
<td>Binary choice, text completion, object identification and picture description tests</td>
<td>Prompts &gt; recasts &gt; control on written measures</td>
</tr>
<tr>
<td>Ammar &amp; Spada (2006)</td>
<td>Sixth graders learning ESL</td>
<td>possessive determiners: his and her</td>
<td>11 communicative sessions</td>
<td>Passage correction and picture description</td>
<td>Recasts &lt; prompts for low-proficiency learners; Recasts = prompts for high-proficiency learners</td>
</tr>
<tr>
<td>Ellis, Loewen &amp; Erlam (2006)</td>
<td>Low-intermediate ESL private school students</td>
<td>Past tense -ed</td>
<td>2 storytelling sessions (1 hour in total)</td>
<td>Elicited imitation, untimed GJT and metalinguistic tests</td>
<td>Recasts &gt; control (oral imitation delayed); Recasts &lt; repetition-metalinguistic (oral imitation and untimed GJT delayed)</td>
</tr>
<tr>
<td>R. Ellis (2007)</td>
<td>Adult ESL private school students</td>
<td>Past tense -ed &amp; comparative -er</td>
<td>2 storytelling sessions for -ed and description activities for -er</td>
<td>Elicited imitation, untimed GJT, and metalinguistic tests</td>
<td>Recasts = control for both targets; Recasts &lt; (repetition-)metalinguistic</td>
</tr>
<tr>
<td>Sheen (2007)</td>
<td>Adult ESL learners</td>
<td>English articles</td>
<td>2 sessions of narrative task</td>
<td>Speeded dictation, writing and error correction tests</td>
<td>Recasts &lt; correction-metalinguistic; Recasts = control</td>
</tr>
<tr>
<td>Ammar (2008)</td>
<td>Sixth graders learning ESL</td>
<td>possessive determiners: his and her</td>
<td>11 communicative sessions</td>
<td>Computerized fill-in-the-blank and picture description tests</td>
<td>Prompts &gt; recasts &gt; control for low level learners</td>
</tr>
<tr>
<td>Yang &amp; Lyster (2010)</td>
<td>University EFL learners</td>
<td>English past tense forms</td>
<td>2 dictogloss, 1 question-and-answer and 1 picture-cued-narrative tasks</td>
<td>Oral and written production tasks</td>
<td>Recasts &lt; prompts for regular past; Recasts = prompts for irregular past.</td>
</tr>
<tr>
<td>Goo (2012)</td>
<td>University</td>
<td>English that-</td>
<td>2 one-way information</td>
<td>GJT and written</td>
<td>Recasts &gt; control</td>
</tr>
<tr>
<td>Mifka-Profozic (2013)</td>
<td>intermediate EFL learners</td>
<td>trace filter Passé composé; imparfait</td>
<td>gap tasks 3 picture-based narrative tasks</td>
<td>production tests Oral and written versions of picture-based narrative tasks</td>
<td>Recasts = metalinguistic Recasts &gt; clarification requests on oral &amp; written tasks; Clarification requests&gt;control only in the written delayed posttest for imparfait</td>
</tr>
</tbody>
</table>
These classroom-based studies have reported mixed results regarding the efficacy of recasts: less effective than (Ammar, 2008; Ammar & Spada, 2006; R. Ellis, 2007; Ellis et al., 2006; Lyster, 2004; Sheen, 2007; Yang & Lyster, 2010), equally effective as (Goo, 2012; Yang & Lyster, 2010), or more effective (Mifka-Profozic, 2013) than a particular type or a combination of prompts. Recasts were not superior to the control condition in R. Ellis (2007) and Sheen (2007).

Factors accounting for the conflicting results may be treatment length, CF manipulation, target structures, choice of measures of linguistic knowledge, and learner proficiency. R. Ellis (2007) and Sheen’s (2007) treatments both consisted of two sessions, and the two sessions lasted for about one hour in total. The comparatively short treatment may have partially constrained the utility of recasts in their studies.

There is an important methodological issue that needs to be considered when comparing recasts and prompts. Prompts are an umbrella category covering both implicit and explicit CF. As Ellis and Shintani (2014) argued when referring to Mifka-Profozic (2013), it is possible that the explicitness of prompts contributes to their superiority rather than self-correction that they elicit. The implicit prompt (i.e. clarification request) in Mifka-Profozic (2013) was found to be less effective than recasts. Even an explicit type of prompt (metalinguistic CF) may not be necessarily more beneficial than recasts when implemented as a single move, as shown in Goo (2012). It is likely that recasts are as effective as a single type of prompt, and even more facilitative of L2 learning than implicit prompts.

Target structures may partially be responsible for the mixed results for recasts. In Yang and Lyster’s (2010) study, there was no difference between CF types in learning irregular past forms but there was a larger effect for prompts than recasts in learning regular past forms. The effect of metalinguistic CF in R. Ellis (2007) was evident in the immediate posttest for comparative -er but in the delayed posttest for past tense -ed. Recasts in R. Ellis’ study did not work for either of the target features.

The relative efficacy of CF can be examined using different measures of linguistic knowledge. In Lyster’s (2004) study prompts were more effective than recasts in the written tests but not in the oral tests. However, this comparative effect was found in the oral tests in Ammar and Spada (2006). To examine the effect of CF on implicit and explicit knowledge, Ellis et al. (2006) used oral imitation, untimed GJT and metalinguistic tests. It was found that
metalinguistic CF outperformed recasts in both implicit knowledge (oral imitation) and explicit knowledge (GJT) tests. However, Mifka-Profozic (2013) reported a greater effect for recasts than for implicit prompts in both oral and written versions of narrative tasks. The reported effect of CF may depend on what language tests were employed. It will be necessary to include a variety of language tests in future investigation of the effectiveness of recasts in order to establish what effect recasting has on declarative/explicit and procedural/implicit knowledge.

As for learner proficiency, high-proficiency learners benefit from recasts as well as prompts. That is to say, low proficiency may prevent the learners from benefiting from recasts. This claim is supported by Ammar and Spada (2006) and Ammar’s (2008) finding that for learners with little prior knowledge of the target structure, recasts are less effective than prompts. Recasts are more beneficial than implicit prompts when learners have high levels of prior knowledge, as shown in Mifka-Profozic (2013).

To date there is no empirical evidence supporting the claim that recasts may be more effective in facilitating the acquisition of new features. In the previous classroom-based studies investigating recasts, the learners invariably had some prior knowledge of the target structures. Learners with greater prior knowledge benefited from recasts as well as prompts. For learners with little prior knowledge, recasts were comparatively ineffective in classroom situations. These findings seem to suggest that recasts work better for partially acquired forms than for new forms. Future research needs to choose target structures that the learners lack prior knowledge of but are developmentally ready to learn, so that the effect of recasts on new linguistic forms can be investigated.

### 3.6 Meta-analyses of corrective feedback

Meta-analyses synthesize the results of experimental and quasi-experimental studies in a specific research area by taking a quantitative approach (i.e. calculating Cohen’s (1988) $d$ index). Meta-analyses in SLA have been conducted to address the effectiveness of oral interaction, corrective feedback, or recasts in particular and moderator variables.

Russell and Spada (2006) selected 15 from 56 primary studies of oral and written feedback on grammar learning, finding that CF in general is effective with a mean effect size of $d = 1.16$ for the treatment-control comparison in the immediate posttest, which is quite large, and $d = .98$ in the delayed posttest, which indicates the durability of CF. Due to the small sample
size, no robust evidence can be found for the differences in the type of CF, mode of CF ($d = .91$ for oral and $1.31$ for written CF), focus of CF, and settings.

Mackey and Goo (2007) carried out a synthesis of 28 studies of the effect of interaction on the development of lexis and grammar. Their findings confirmed the immediate ($d = .75$), short-term ($d = 1.02$) and long-term ($d = .99$) effects of interaction, though these results should be interpreted with caution because there were only eight primary studies using a longer-term delayed posttest. Interaction is more effective for lexis in the short term, but for grammar in the long term. The effect size of recasts was found to be large ($d = .96$ in immediate, $1.69$ in short-term delayed and $1.22$ in long-term delayed posttests) but no firm conclusion was reached on whether recasts were superior to other types of feedback given the small number of comparative studies. The effectiveness of interaction varied depending on context (foreign > second language), setting (laboratory > classroom), tests of learning outcome (closed-ended prompted production > open-ended prompted production or prompted response).

Li (2010) synthesized 33 studies of oral corrective feedback as an isolated construct. A medium and durable effect size was found for CF in general ($d = .61$ in immediate, $.57$ in short-term delayed and $.54$ in long-term delayed posttests) and recasts in particular ($d = .51$ in immediate, $.44$ in short-term delayed and $.53$ in long-term delayed posttests). The effect of implicit CF involving recasts was more durable than that of explicit CF though the latter had a greater immediate effect. The length of treatment (long > short), settings (laboratory > classroom) and contexts (foreign > second language) also affected the efficacy of CF.

Due to the difference in effect size between laboratory- and classroom-based studies, Lyster and Saito (2010) focused on 15 classroom quasi-experimental studies of oral feedback to examine the CF’s pedagogical value. They found that the overall effect of oral CF was medium-to-large ($d = .74$) and the effect of recasts was medium ($d = .53$), smaller than that of other types of CF ($d = .83$ for prompts and .84 for explicit correction) in classroom settings. The impact of CF also differed across language tests (large ES for free constructed responses), treatment length (larger ES for long and brief than short treatment) and learner age (larger ES for young learners) but not between contexts (foreign vs second language).

Miller and Pan’s (2012) meta-analysis of 17 publications focused on recasts in L2 classrooms, reporting a “modestly significant” mean effect size ($d = .38$). Although the effect size was
weighted due to the heterogeneity across primary studies, 40 effect sizes for treatment retrieved from 17 publications potentially posed a risk because retrieval of more than one effect size from a primary study would render the sample size inflated (Li, 2010). The effect of recasts on language learning was found to vary depending on the type of treatment, the nature of grammatical structure and the L1-L2 difference.

Table 6: Meta-Analyses of CF Involving Recasts

<table>
<thead>
<tr>
<th>Meta-analysis</th>
<th>Number of primary studies</th>
<th>Research focus</th>
<th>Effect size (d)</th>
<th>Moderating factors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Russell &amp; Spada (2006)</td>
<td>15</td>
<td>Oral and written CF</td>
<td>.91, 1.31</td>
<td>No evidence</td>
</tr>
<tr>
<td>Mackey &amp; Goo (2007)</td>
<td>28</td>
<td>interaction</td>
<td>.96, 1.69, 1.22</td>
<td>Linguistic focus, context, setting and language tests</td>
</tr>
<tr>
<td>Li (2010)</td>
<td>33</td>
<td>Oral CF</td>
<td>.51, .44, .53</td>
<td>Treatment length, setting and context</td>
</tr>
<tr>
<td>Lyster &amp; Saito (2010)</td>
<td>15</td>
<td>Classroom oral CF</td>
<td>.53</td>
<td>Language tests, treatment length and age</td>
</tr>
<tr>
<td>Miller &amp; Pan (2012)</td>
<td>17</td>
<td>Classroom recasts</td>
<td>.38</td>
<td>Treatment type, linguistic focus and target language</td>
</tr>
</tbody>
</table>

As Table 6 shows, meta-analyses have revealed that interaction with CF facilitates second language learning and its facilitative role seems to be influenced by linguistic focus, target language, treatment length, treatment type, context, testing methods and learner age. These meta-analyses confirm the positive effect for interaction and CF on L2 learning, but the results for recasts are quite mixed. When recasts are addressed as a sole construct, their effect size is reported to be large (e.g. Mackey & Goo), medium (e.g. Li) and small (e.g. Miller & Pan).

The small effect size reported for classroom recasts (Miller & Pan) indicates that recasts are possibly more susceptible to the impact of various contextual variables in real classrooms. It is worthwhile to conduct more classroom-based research on recasts. The discrepancy on the efficacy of recasts between laboratory and classroom settings raises the issue of “ecological validity”, which Ellis et al. (2006) argued can only be achieved by doing classroom-based
research. It may not be possible to single out the most effective CF given the presence of various factors in classroom settings (Lyster et al., 2013) and the possibility is that different types of CF work for different types of linguistic knowledge – recasts facilitate the learning of new forms while prompts enhance the control of existing forms (Ellis & Shintani, 2014). Goo and Mackey (2013) argued that future research should not compare “apples” and “oranges” but focus on recasts to see under what conditions this type of CF works best for L2 learning. For instance, does explicitness make a difference to the efficacy of recasts and does working memory capacity mediate the effectiveness of recasts?

3.7 Recasts and Individual differences in working memory

Working memory is a cognitive construct that temporarily stores and processes the information required to perform complex cognitive tasks. Robinson (2005) posited that WM capacity, in combination with other cognitive abilities, determines the extent to which recasts benefit learners. N. Ellis (2005) also proposed that not only phonological short term memory (PSTM) but also CWM capacity is likely to affect the extent to which recasts are noticed and retained, and thus mediates the efficacy of recasts. To date, only a few studies have empirically explored the relationship between WM, recasts and L2 learning.

Mackey et al. (2002) examined the link between composite scores for listening span and nonword span and recast-driven language development after NS-NNS interaction, claiming that WM was potentially related to the efficacy of CF, but in different ways. Given the small number of participants, they tentatively concluded that low-WM learners achieved more substantial gains from recasts immediately after the treatment, but high-WM learners had greater development in the long term. Sagarrá (2007) confirmed the delayed effect of WM, but argued for a positive relationship between WM (measured by a reading span test) and the immediate effect of recasts, because high-WM learners outperformed low-WM learners in learning Spanish noun-adjective agreement after receiving computer-delivered recasts. Goo (2012) found that not only reading-span but also operation-span WM scores predicted the effect of recasts, emphasizing that WM, especially executive attention, was involved in noticing this comparatively implicit type of feedback. Révéz (2012) examined the ways that WM mediated the efficacy of recasts in different tests of linguistic knowledge. It was found that reading span capacity was positively related to written achievement. However, Mifka-Profozic (2013) reported that the correlation of WM and immediate gains for passé composé driven by recasts was positive in oral tests but negative in written tests.
In contrast, other researchers reported a weak relationship between WM and the utility of recasts. For instance, Li (2013a) found that neither analytic ability nor CWM predicted the effect of recasts on the learning of Chinese perfective -le in elicited imitation and grammaticality judgement tests. Trofimovich et al. (2007) detected a very weak correlation between CWM and recast-driven development. They inferred that the discrepancy with previous research might be due to their different WM measures (serial nonword recognition and letter-number sequencing) and their operationalization of recasts. The salience and predictability of their recasts rendered noticing of corrective feedback less demanding and thus minimized the role of individual differences in WM. However, the possibility that noticing of explicit recasts is not sensitive to WM needs further investigation.

As for the role of phonological short-term memory (PSTM) in processing recasts, results are also unclear. Among the above studies, Mackey et al. (2002) found PSTM scores were negatively correlated with initial language development, but positively with long-term development. In Trofimovich et al. (2007), PSTM predicted “delayed posttest” linguistic accuracy, but in fact this test was administered only 2 to 12 minutes after the immediate posttest. Révész (2012) also showed a significant positive correlation between PSTM and recasts-driven immediate development in oral tests.

I will now consider the role of PSTM and CWM in recasts-driven L2 learning. Taken together, these studies indicate that both types of working memory are generally related to L2 learning in the long run, but questions remain about their contribution to the short-term effect of recasts. There are still inconsistent findings about PSTM: positively correlated with the immediate effect of recasts in Trofimovich et al. (2007) and Révész (2012) but negatively in Mackey et al. (2002). As for the role of CWM, mixed findings also exist. There seems to be no relationship between WM and the effect of recasts in Li (2013a) and Trofimovich et al. (2007). Those studies arguing for the role of CWM in recasts also differ: positively associated with written production in Révész (2012) but negatively in Mifka-Profozic (2013). Mifka-Profozic’s (2013) results seem to suggest that CWM mediates the effect of recasts on implicit knowledge as measured by oral tasks but not on explicit knowledge as measured by written tasks. However, Révész’s (2012) findings indicate that CWM contributes more to the acquisition of declarative knowledge whereas PSTM is important for the proceduralization of existing knowledge. How WM components and knowledge types interact in mediating the
effectiveness of recasts constitutes a new area of enquiry, which necessitates the use of different tests to measure WM and implicit and explicit linguistic knowledge.

### 3.8 Conclusion

The role of recasts in L2 learning has been investigated in terms of uptake, noticing and acquisition. Recasts are widely used as an error correction technique in various settings – meaning-oriented and form-oriented, second-language and foreign-language classrooms – but tend to generate fewer instances of uptake than prompts. What may constrain uptake rate following recasts may be the nature of the CF, the type of errors, opportunities for response, learner factors and contextual variables. However, it remains unclear whether learner factors such as age and contextual variables such as learners’ orientation moderate the level of successful uptake.

No consensus has been reached regarding the role of uptake in L2 acquisition. The few studies that have investigated both uptake and learning outcomes suggest that uptake of recasts involves mechanical repetition and may not be related to acquisition, but Révész et al. (2011) did report a potential link in their simple task condition. Future research needs to consider factors such as the nature of tasks, the explicitness of recasts, the types of uptake (repetition or primed output), and the tests of linguistic knowledge in investigating the potential uptake-learning relationship.

It is likely that recasts contribute to acquisition not via the uptake they lead to but the noticing of target language forms that they trigger. Laboratory-based studies report that recasts enhance noticing. Classroom studies show that explicit recasts are easier to attend to than implicit recasts. The noticeability of implicit recasts is likely to be affected by the salience of the target structure, learner proficiency, learner form-orientation and working memory capacity as demonstrated by laboratory-based studies. Future research on how these variables mediate the noticeability of recasts needs to use a variety of methods investigating noticing.

Studies of recasts and acquisition using posttests indicate that explicit recasts drive L2 learning but the effect of implicit recasts remains controversial. It is not clear whether recasts facilitate acquisition due to negative evidence in addition to the positive evidence they provide. To resolve this uncertainty, future research needs to compare explicit and implicit recasts. The two comparative studies that have investigated this reported conflicting results:
explicit recasts are more effective than implicit recasts in Chen (2010) but no difference was reported in Erlam and Loewen (2010).

Comparative studies of recasts, prompts and models in laboratory settings provide evidence of a positive effect for recasts. However, the relative effect of recasts was evident for some target structures but not for others, and in some tests of linguistic knowledge but not in others, indicating the necessity to consider the impact of these factors on the utility of recasts in future research.

Comparative studies in classroom settings have produced mixed results. Recasts have been found to be less effective than prompts in some studies but more effective in others and sometimes not effective at all (i.e. in comparison to a control condition). What mediates the relative effect of recasts may be treatment length, CF manipulation, choice of target structures, choice of linguistic tests, learner individual differences and contextual variables. Meta-analyses confirm the mediation of these factors on the efficacy of interaction with CF but controversies remain regarding how they mediate the effect of recasts, for instance, the effect of the length of treatment. Explicit recasts lead to more repair than implicit recasts, but whether explicitness in recasts makes a difference to learning outcomes needs to be verified. The effect of recasts varies according to the target structure, but it remains unclear whether recasts promote more development of new or existing linguistic forms. It is also necessary to consider the influence of the choice of language tests. We need to investigate whether recasts facilitate the development of implicit, explicit or both types of linguistic knowledge. Recasts seem to benefit high-proficiency learners to a greater extent than low-proficiency learners, but whether other learner factors mediate the effectiveness of recasts warrants further study.

Only a few studies have investigated the role that WM plays in processing recasts and no firm conclusions are possible. Some studies failed to find a correlation between CWM and recasts-driven L2 learning, others reported a significant correlation but the results were mixed: WM was sometimes positively and sometimes negatively related to the immediate posttest gains promoted by recasts. The results regarding the potential link between phonological short-term memory and the immediate effect of recasts were also conflicting.

Context and setting are also moderating variables. Laboratory-based studies have reported a positive effect for the CF but classroom-based studies often show mixed results. Recasts are
assumed to be more effective in form-oriented than meaning-oriented classrooms according to the Counterbalance Hypothesis.

The quasi-experimental study reported in this thesis addresses some of the above-mentioned issues regarding recasts in a context where orientation is traditionally on language accuracy. The effects of explicit and implicit recasts are compared on the acquisition of two English structures using tests of implicit and explicit knowledge, aiming to investigate whether recasts contribute to L2 grammar learning and how the target structure and the type of linguistic knowledge test mediate the CF’s efficacy. Correlational analyses of uptake and learning outcomes shed light on whether successful uptake is related to learning. Correlational analyses of WM and gains in implicit and explicit knowledge shed light on the role of WM in processing recasts.
Chapter 4 Methodology

Chapter 4 describes the research methodology for the present study, which uses a quasi-experimental design to test the effects of explicit and implicit recasts and a correlational design to examine the relationship between repair after recasts, acquisition of L2 structures and working memory (WM). It is comprised of research objectives and questions, research design, participants, instruments, and analytical procedures.

4.1 Research objectives and questions

The objectives of this study were to investigate whether recasts facilitated instructed acquisition and whether explicitness played a role in the efficacy of the recasts. The effects of explicit recasts and implicit recasts were compared in terms of implicit knowledge and explicit knowledge. A second objective was to ascertain to what extent learners repaired their errors following corrective and implicit recasts and whether uptake-with-repair was related to subsequent acquisition. A third objective was whether the learners’ working memory capacity (phonological short-term memory and complex working memory) affected their ability to repair in response to recasts and whether it was related to their acquisition of target structures. The final objective was whether there was any difference between the two target structures, 3rd person -s and embedded questions, in uptake rate following recasts, implicit knowledge gains, explicit knowledge gains, the uptake-acquisition relationship, and the WM-acquisition relationship.

The specific research questions (RQ) this study sought to answer are as follows.

RQ1. What effects do corrective recasts and implicit recasts have on learners’ implicit knowledge of L2 grammatical features (3rd person -s and embedded questions)?
   a. What effect do corrective recasts have on learners’ implicit knowledge of L2 grammatical features?
   b. What effect do implicit recasts have on learners’ implicit knowledge of L2 grammatical features?
   c. Is there any difference between the effects of corrective and implicit recasts on learners’ implicit knowledge?

RQ2. What effects do corrective recasts and implicit recasts have on learners’ explicit
knowledge of L2 grammatical features?

a. What effect do corrective recasts have on learners’ explicit knowledge of L2 grammatical features?

b. What effect do implicit recasts have on learners’ explicit knowledge of L2 grammatical features?

c. Is there any difference between the effects of corrective and implicit recasts on learners’ explicit knowledge?

RQ3. Is there any difference in uptake-with-repair following corrective recasts and implicit recasts?

RQ4. Is there any relationship between learners’ uptake-with-repair following recasts and their subsequent acquisition of implicit and explicit knowledge?

RQ5. What role does working memory play in corrective feedback involving recasts?

a. Do differences in learners’ working memory mediate their ability to repair following i) corrective recasts and ii) implicit recasts?

b. Do differences in learners’ working memory mediate their acquisition of target structures i) in the corrective recast condition and ii) in the implicit recast condition?

RQ6. Is there any difference between the two target structures in uptake rate following explicit/implicit recasts, recasts-driven implicit/explicit knowledge development, and the role of working memory in CF involving recasts?

4.2 Research design

To observe the relative effectiveness of corrective recasts and implicit recasts, a quasi-experimental design was adopted, using four groups of English learners from pretest, treatment, immediate posttest, to delayed posttest. This design was repeated as there were two target structures: Study 1 for 3rd person -s and Study 2 for embedded questions (see Figure 1). Also woven into this research was a correlational design to investigate the relationship between uptake, knowledge gains and WM. Uptake data were collected by recording researcher-learner interactions during the treatment. WM data were gathered by administering three independent tests, namely the nonword span test, the digit span test and the listening span test, at the end of the experiment. The former two tests were intended to tap into phonological short-term memory (PSTM) and the last one into complex working memory (CWM). The whole data collection process ended with an exit questionnaire, which aimed to provide evidence for noticing in addition to the evidence from examining uptake.
Figure 1: Research Design

Note. EI=elicited imitation; OP=oral production; WP=written production; UGJ=untimed grammaticality judgment.

All the above sessions were completed within 13 weeks in the following order: Study 1 pretest, Study 1 treatment, Study 1 immediate posttest, Study 2 pretest, Study 2 treatment, Study 2 immediate posttest, Study 1 delayed posttest, working memory tests, Study 2 delayed
As Figure 1 shows, the test control group only took the four language tests (pretest, immediate posttest and delayed posttest): elicited imitation (EI), oral production (OP), written production (WP) and untimed grammaticality judgment (UGJ). The first two tests were assumed to measure implicit knowledge whereas the last two tests were designed to elicit more explicit knowledge. The details of these tests will be presented in Section 4.4.6. The task control group completed task-based lessons as well as the language tests. They performed three focused tasks targeting 3rd person -s and another three tasks for embedded questions, but did not receive any corrective feedback on either structure. The two experimental groups completed the language tests, the tasks for the target structures, working memory tests and an exit questionnaire. When they committed an error in a target structure while performing the tasks, they also received corrective feedback, either corrective recasts or implicit recasts, depending on which group they were in.

4.3 Participants

4.3.1 Learners

After learning what task-based lessons were like, four classes of 1st-year English learners from a northern Chinese university were invited to participate in this research. Two classes majored in English Education and two classes majored in Business English. Each class was evenly divided into two subclasses and then the eight subclasses were paired to form four Education-Business groups for this study: the corrective recasts group (n=33), the implicit recasts group (n=33), the task control group (n=22) and the test control group (n=21). The English Education students accounted for about 54.5% and the Business English students made up 45.5% in each group. The total sample size was 109. However, it should be noted that the sample size varied depending on data analyses as those learners registering a pretest score of over 90 on a specific measure were excluded.

Most of the participants were female, accounting for 94.5% of the total. They were around 20 years old, studying at the English Department of a Chinese university. They had oral English classes in their curriculum, so the majority was able to complete the oral tasks this research required. An EI test administered at the beginning of the study revealed that they already had some procedural knowledge of both 3rd person -s and embedded questions but that they were far from the 90% level of accuracy often used as the criterion for acquisition. For 3rd person -
the pre-test scores for the whole sample was 36.49% while for embedded questions it was 56.23%. Having passed the national college entrance examination, they had most likely acquired quite a large amount of explicit knowledge of the target structures prior to this experiment. This is evident in the pretest scores for the UGJ test. For 3rd person -s the mean score for the whole sample was 69.09% while for embedded questions it was 65.79%. These scores were higher than those on the EI test but still well below the 90% criterion level.

The participants all came from the same province and none of them had ever been abroad so they had had similar English-learning experiences. Their regular teachers implemented an approach involving focus on forms by means of presentation, practice and produce, and occasionally returned to the grammar translation method. A short interview with class teachers after the experiment confirmed that they did not correct errors of 3rd person -s and embedded questions in class during that period of time. None of the participants had any other English training in the meanwhile either.

4.3.2 Researcher

The researcher observed the participants’ routine classes for a week to get familiar with them, introduced task-based language teaching to the students, and invited them to participate in this study. The task-based lessons were taught in their spare time. The researcher worked as a teacher in the corrective recast, implicit recast and task control groups, giving the treatment accordingly in each group. He aimed to correct every 3rd person -s / embedded question error the participants made with either corrective or implicit recasts when it was possible and appropriate to do so in the experimental groups. He recorded these task-based lessons to identify interaction episodes for data analysis.

4.3.3 Assistants

A research assistant helped to organize and invigilate the tests as a reliever. The elicited oral imitation, oral production, nonword span, digit span and listening span tests were all conducted one-on-one, and thus were time-consuming. Whenever a participant was found not following the procedure, this was rectified immediately to ensure test validity.

Two Information Technology students from the same university designed a computer-based exit questionnaire. The questionnaire was computerized in order to encourage the participants to comment on the project freely and to control the order in which they answered the questions so that their responses to later items would not influence their answers to earlier
ones.

4.4 Instruments

In order to collect sufficient data for the present research, three focused tasks for 3rd person -s and three focused tasks for embedded questions were designed together with procedures for corrective recasts (explicit) and implicit recasts. Implicit knowledge tests were operationalized as the elicited oral imitation and oral production tests; and explicit knowledge tests were operationalized as the written production and untimed grammaticality judgment tests. Phonological short-term memory was measured using nonword span and digit span tests; and complex working memory was measured by means of a listening span test. A computerized exit questionnaire was administered at the end of the project.

4.4.1 Target structures

The first structure this research targeted was 3rd person -s. This is a straightforward grammatical rule, that is, every regular verb after a 3rd person singular subject in the simple present tense takes the form of -s (-es) for subject-verb agreement. It is a simple structure in terms of explicit knowledge. R. Ellis (2006a) considered 3rd person -s a typical example of a linguistic structure that “may be inherently easy to learn as explicit knowledge but difficult to acquire as implicit knowledge” (p. 432).

Third person singular is acquired late (see Krashen’s (1982) Natural Order Hypothesis). One possible explanation for its late acquisition is its lack of salience. Salience, that is, how prominent a structure is to hear, is believed to influence the acquisition of target features (Goldschneider & DeKeyser, 2005). Although it has three phonological variants: /s/, /z/ and /iz/, this phoneme is generally hard to perceive, and thus is difficult to acquire as implicit knowledge.

The second target structure is embedded questions, that is, reported questions. They are noun clauses that begin with who, what, where, when, why, how, if, etc. and that require declarative word order. According to Pienemann’s (2005) processability hierarchy, inversion in English wh-questions is learned at Stage 4 and “cancel inversion” in embedded questions occurs at Stage 6, the highest stage. Cancel inversion involves distinguishing between main and subordinate clauses and then the application of declarative order in the subordinate clause. Such a syntactical feature is difficult to acquire as implicit knowledge. However, in some respects embedded questions are arguably easy to acquire because unlike interrogatives there
is no auxiliary verb (except for negative indirect questions). In fact, embedded questions use a canonical word order and this is less likely to constitute a challenge. Thus, there are opposing arguments regarding the acquirability of this structure.

Embedded questions are a syntactical feature whereas 3rd person -s is a morphological feature. The former is comparatively more complex as explicit knowledge but more salient as implicit knowledge than the latter. Therefore, it is interesting to compare the effect of recasts on these two different structures.

4.4.2 Focused tasks for 3rd person -s

Study 1 targeted 3rd person -s, which refers to the verb form in the simple present tense when the subject is 3rd person singular. To create contexts in which a 3rd person singular subject, the simple present tense and action verbs are obligatory, Study 1 used the following 3 focused tasks: A Celebrity, My Partner, and A Chinese City.

Task 1 “A Celebrity” involved daily activities of a famous person. It required the participants to choose a picture and describe what the person in the picture looks like and what he or she usually does for the teacher and other participants to guess who he/she is. Task 2 “My Partner” required the participants to describe and compare their partners’ life style in order to find out whose college life differs most from others’. Task 3 “A Chinese City” required the participants to collect and present information about a Chinese city in order to profile the features of a number of Chinese cities.

These one-way information gap tasks were designed according to the six defining characteristics of an instructional task, that is, “[it] is a workplan”, “involves a primary focus on meaning”, “involves real-world processes of language use”, “can involve any of the four language skills”, “engages cognitive processes” and “has a clearly defined communicative outcome” (R. Ellis, 2003, pp. 9-10). Unfocused tasks allow “learners to choose from a range of forms” whereas “focused tasks aim to induce learners to process, receptively or productively, some particular feature” (p. 16), for instance, 3rd person singular -s. The instructional tasks in this research were designed in a way that they can be completed properly only by using the target structure, that is to say, they constitute focused tasks. The materials and procedures for each instructional task are described below.
Task 1: A celebrity

Materials:

“Use the given words to make sentences saying what the person shown in the picture you chose does. Your sentences should help the teacher and other students guess who the person is.”
<table>
<thead>
<tr>
<th>Celebrity 1</th>
<th>Celebrity 2</th>
<th>Celebrity 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jay Chou</td>
<td>Laure Shang</td>
<td>Chris Lee</td>
</tr>
<tr>
<td>Words:</td>
<td>sing, compose, release, perform, imitate, write, love, act, whirl, translate, look like, speak, wear, dance and sing, hold, ask, raise, lead, work with / for</td>
<td></td>
</tr>
<tr>
<td>Guo Jingjing</td>
<td>Yi Jianlian</td>
<td>Yao Ming</td>
</tr>
<tr>
<td>Words:</td>
<td>dive, win, engage, suffer from, say, appear, keep, play, speak, work with / for, shoot, impress, wear, dunk, enroll, take, endorse, participate, raise, donate, invest</td>
<td></td>
</tr>
<tr>
<td>Xiao Shenyang</td>
<td>Zhao Benshan</td>
<td>Li Yong</td>
</tr>
<tr>
<td>Words:</td>
<td>perform, act, show off, imitate, wear, utter, amuse, direct, highlight, play, feel, ridicule, make, create, bring, own, win, work for, live, host, laugh, run into, stretch out, make</td>
<td></td>
</tr>
<tr>
<td>Jackie Chan</td>
<td>Jet Li</td>
<td>Stephen Chow</td>
</tr>
<tr>
<td>Words:</td>
<td>perform, sing, direct / shoot, act / star / play, produce, appear, release, hold, break, contribute / donate, support, own, speak, wear, found, like, ride, read, meditate, collect, believe in, smile</td>
<td></td>
</tr>
<tr>
<td>Feng Xiaogang</td>
<td>Jiang Wen</td>
<td>Zhang Yimou</td>
</tr>
<tr>
<td>Words:</td>
<td>direct, specialize in, use, express, consider, act / star / play, write, impress, produce, win, explore, cast / feature, object to, seek, adapt, look, (seldom) talk, insist, wear</td>
<td></td>
</tr>
<tr>
<td>Zhou Libo</td>
<td>Guo Degang</td>
<td>Feng Gong</td>
</tr>
<tr>
<td>Words:</td>
<td>create, base, deliver, comment on, perform, (never) script, speak out, wear, work for, achieve, act, hope, influence, quarrel with, revitalize, own, direct, write, appear, help, win, take up, say, see</td>
<td></td>
</tr>
<tr>
<td>Yun-fat Chow</td>
<td>Chen Daoming</td>
<td>Ge You</td>
</tr>
<tr>
<td>Words:</td>
<td>act / play / star, win, move, wish, appear, care, talk, enjoy, look, adapt to, begin, read, practice, (never) drink, keep, wear, live, laugh at, behave, work as / endorse, impress</td>
<td></td>
</tr>
</tbody>
</table>

**Figure 2: Celebrity Pictures**
**Procedures:**

- Students worked in 7 groups. Each group was given a set of three pictures of famous people. It had to choose ONE of these pictures and work together to make sentences about this person using the words listed beneath the pictures (e.g. what he looks like, what he usually does, and what hobbies he has).

- The teacher explained that they should only say what the person DOES rather than who he / she IS, and that they should NOT mention the singer of a song, the director of a film, the founder of an organization or any personal name, in order to make it more difficult to guess.

- The teacher selected a leader for each group. The leaders took turns to tell the teacher and the rest of the class about the person in the picture their group had chosen.

- When necessary, the teacher asked questions to find out more information about the person in the picture (e.g. “Where does he live?”, “What does his wife do”, and “Is there anything else you know about him?”).

- The teacher recast those participants’ utterances that contained a 3rd person -s error.

- When the teacher was sure who the person in the picture was, he asked the students to write down the name of the person. He then wrote the name on the board. Students who got the correct name got one point.

- Then, another member in each group was appointed “leader” and these new leaders completed the 2nd round of tasks.

- After all the people in the selected pictures had been described, each student in the class added up the number of names they had got correct. The student with the most correct names was the winner.

**Task 2: My partner**

**Materials:**

“Work with a partner. Ask your partner questions to enable you to complete a table about his/her life style. Fill in the table with the information you obtained by asking him/her. Your partner will then ask you questions and fill in the table about you.”
Table 7: My Partner’s Life Style

<table>
<thead>
<tr>
<th>Time</th>
<th>Activities</th>
<th>Further Information</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Get up</td>
<td>Regularly?</td>
</tr>
<tr>
<td></td>
<td>Have classes</td>
<td>What? Like or skip?</td>
</tr>
<tr>
<td>In the evening</td>
<td>Do physical exercises / sports</td>
<td>What? How often?</td>
</tr>
<tr>
<td>At weekends</td>
<td>Engage in social activities</td>
<td>What? How?</td>
</tr>
<tr>
<td></td>
<td>Have hobbies</td>
<td>What? How often?</td>
</tr>
</tbody>
</table>

“When you are listening to a speaker describing his/her partner’s life style, please answer these questions. 1) Is the description you are listening to similar to or different from your own partner’s life style? 2) What are the differences?”

Procedures:

- When the participants had completed the tables in pairs, the teacher invited one individual student in each pair to come to the front of the class and tell the teacher and the rest of the class about his/her partner’s life style.

- The teacher listened, commented and asked for more information about the partner’s life style.

- The teacher recast any utterance containing a 3rd person -s error.

- As students listened, they took notes to decide whether the description they heard is similar to or different from their own partner’s life style and what the differences are.

- The teacher invited a listener to report what he/she found in the description that is different from his/her own partner’s life style.

- The teacher took this opportunity to recast any utterance containing a 3rd person -s error.

- Finally, the teacher listed each student’s unique characteristics as mentioned during the session.

Task 3: A Chinese city

Materials:
“Work in small groups. Each group collects and presents information about one of seven cities in China: Beijing, Shanghai, Guangzhou, Xi’an, Hangzhou, Lhasa and Shijiazhuang. We are going to explore facts about each city relating to its geography, climate, transportation, attractions and unique features, and finally we will produce a profile of each city.”

| Table 8: Facts of Beijing/Shanghai/Guangzhou/Xi'an/Hangzhou/Lhasa/Shijiazhuang |
|-------------------------------------------------|---------------------------------|
| **Basic Facts**                                 | **Useful Words**                |
| Geography                                       | cover, adjoin, equal, lie, encompass, consist of, border, run through, flow, form |
| Climate                                         | enjoy, go up to, last, increase, blow, vary, occur, come, decrease, change |
| Transportation                                  | take, service, offer, connect, remain, enable, become, run, consist of |
| Attractions                                     | remain, represent / reflect, show, boast, contain, attract / draw |
| Unique Features                                 | provide, emphasize, serve as, show off, draw, hold, possess, date back |

**Procedures:**

- Students worked in seven groups. Each group collected information about basic facts about a Chinese city prior to the task session.
- Group leaders were appointed at the outset of the task performance. The leaders led a group discussion on how to present the information using the given words.
- The leaders took turns to come to the front of the class and reported the basic facts they knew about the city.
- The teacher listened and asked for more information about the city. He recast any utterance containing a 3rd person -s error.
- The other groups listened carefully and took notes. When they missed a point, they were allowed to ask the speaker for clarification.
- The teacher asked some of the students (especially those who had not interacted with the teacher yet) to report what they had learned about one of the seven Chinese cities.
- The teacher took this opportunity to recast any utterance containing a 3rd person -s error.
- Finally, the teacher and the participants worked together to produce a profile for each city on the basis of their notes.
4.4.3 Focused tasks for embedded questions

Study 2 targeted embedded questions. The embedded question tasks were designed to create contexts which required the participants to talk about what questions they had asked or they wanted to ask. Task 4 “Jackie Chan Interview” required the participants to fill in missing questions in an interview transcript and then report the questions from memory. Other participants checked and pointed out any questions he or she failed to report. Task 5 “Eating Habits Survey” required the participants to develop a survey questionnaire in groups and then report their specific questions to the class. They had to justify each question they proposed in order to gain a score. Task 6 “Question-raising Competition” aimed to find out typical questions this generation of young people would like to ask their parents, friends, idols, scientists and government. When a participant put forward and justified a question, the teacher wrote the question on the board for other participants to vote for or object to. All these focused tasks had a clearly defined communicative outcome, that is, a complete interview with Jackie Chan, a complete questionnaire on eating habits and a list of typical questions they were interested in.

Task 4: Jackie Chan interview

Materials:

“Work in seven small groups. Imagine you are interviewing Jackie Chan. Complete the interview by writing in the questions you asked Jackie.”
An Interview with Jackie Chan

You: ?
JC: I’m working on a movie called “The Kung Fu Kid” with the son of Will Smith, Jaden Smith. Basically, it’s a remake of “The Karate Kid”, but totally different story, a young kid from America coming to China, and I take care of him. I teach him kung fu. I teach him attitude. Teach him discipline. It’s a very interesting story.

You: ?
JC: Yes, since the 70s and 80s, I’ve been doing action comedy and stunts for decades.

You: ?
JC: I like action but I hate violence. But the only thing I like is comedy with action. I don’t know why. Maybe because I’ve been training my martial arts for so many years and also, I like to make people smile. I like to make the audience excited.

You: ?
JC: I don’t think the audience like watching Jackie Chan running around Krabi beach, kissing and singing a song ... The audience always says “go and see Jackie Chan movie, very funny and good stunts and good actions”, nobody says “go and watch, good crying”. It’s different, so, I think they like my action comedy.

You: ?
JC: I think the audience in Asia has become more Westernized. American movies are very strong all these years, not only movies, but Internet, TV drama, or rap, song. Everybody is learning American culture.

You: ?
JC: It’s a good thing, but shouldn’t be only one side, should be many sides. When you wear a dress like Americans, I don’t even know if you are a Chinese, Japanese, or Korean, but if you do have a sign, let me see, what kind of person, then I do hello. Then when you see Japanese, you go oh-hi-o. You know, your culture.

You: ?
JC: The problem is, piracy is so strong and the investors don’t get back what they want. They spend 50 million to invest in a movie and the box office is only like 5, 6 million.

You: ?
JC: Big problem for everybody. They have fake Gucci bag, fake Cartier, fake everything and then after, I find out fake things don’t matter to the movies, the important thing is they make fake aeroplane accessories, and it makes aeroplane crash, that’s a big problem. How can you make fake engines and medicines, how many people you kill, they don’t know.

You: ?
JC: First, you have to learn so many things as an actor. I always tell my son, don’t only be a singer. You learn to be a musician, piano, guitar, composer, write and sing, then you can live longer. As an action star, you don’t only know how to fight, you must know how to direct, how to write your own script, know the camera angle, know the editing, and read and see more things, accept the world culture, then you can make the movie international. Yes, some stars are just local stars, good, but if you want to be international stars, eyes wider, see more things, accept more things in your mind.

You: ?
JC: Me myself, none. I always think the next movie will be better.
Procedures:

- The participants worked in seven groups to complete the interview by writing in their questions and then they memorized the questions.

- When all the groups had completed their interviews, group leaders were selected in turn to tell from memory what questions their group asked and to explain why they asked such a question. The speaker was not allowed to refer to a copy of the completed interview.

- Each group leader was only given four minutes to report and comment on the questions.

- When a group leader in the corrective recast group reported a direct question (e.g. “How did you become an international star?”), the teacher first responded with an elicitation (e.g. “You asked him how…” or “You want to know how…”). When the student said “I asked him/I want to know how did he/you become an international star”, the teacher repeated and then recast it. However, the teacher simply recast the student’s original utterance with no elicitation and no repetition in the implicit recast group.

- When the student raised a question that did not match Jackie Chan’s responses in the interview, the teacher negotiated with him/her about the question.

- The rest of the class listened carefully and identified what questions the group leader missed. Then they were invited to tell the teacher any question that had not been mentioned.

- When these students committed embedded question errors, the teacher recast their erroneous utterances.

- The teacher made a note of the number of questions each group leader missed. The group whose leader missed the fewest questions won the game.

Task 5: Eating habits survey

This task involved developing a survey questionnaire, reporting specific questions and justifying each question proposed.

Materials:

“Work in small groups. Complete the plan for a survey of eating habits by writing in specific
questions that your group wants to ask. Your questions should serve the general purpose of learning about local people’s eating habits.”

### Table 10: My Plan for an Eating Habits Survey

<table>
<thead>
<tr>
<th>Topic</th>
<th>Eating habits</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>General Question</strong></td>
<td>What eating habits do local people have?</td>
</tr>
<tr>
<td></td>
<td>e.g. How often do you eat green vegetables?</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><strong>Specific Questions</strong></td>
<td><em>(The exact questions that your group will ask people)</em></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>
</tbody>
</table>

**Procedures:**

- The participants worked in seven groups. Each group discussed what specific questions should go into its survey plan.

- After all the groups had completed the survey plan, group leaders came to the front of the class and took turns to report what questions their group had planned to ask. They reported the questions from memory.
• When a group leader presented a question, the teacher commented on it and sometimes asked him/her to explain the purpose of the question.

• When a group leader in the corrective recast group reported a direct question (e.g. “What do you eat for breakfast?”), the teacher first responded with an elicitation (“You want to ask people what ...”). When the student said “We want to ask people what do they/you eat for breakfast”, the teacher repeated and then recast it. However, the teacher simply recast the student’s original utterance with no elicitation and no repetition in the implicit recast group.

• The rest of the class listened carefully to the question each group leader was reporting and decided whether the question was necessary. They were encouraged to dispute any question they considered unnecessary.

• The teacher made a note of how many reasonable objections each group leader encountered. The group whose leader received the fewest objections won the game.

• All the groups improved their survey plan after class.

Task 6: Question-raising competition

Task 6 involved discussing, presenting, justifying, disputing and selecting typical questions that college students would like to ask.

Materials:

“Discuss with your group members what questions you want to ask your parents, friends, idols, scientists and government most and then decide which questions your group will present at the question-raising competition. Write the questions in the given table and memorize them for the competition.”

Table 11: Questions We Want to Ask Most

<table>
<thead>
<tr>
<th>Questions</th>
</tr>
</thead>
<tbody>
<tr>
<td>To our parents</td>
</tr>
<tr>
<td>To our friends</td>
</tr>
<tr>
<td>To our idols</td>
</tr>
<tr>
<td>To our scientists</td>
</tr>
<tr>
<td>To our government</td>
</tr>
</tbody>
</table>
Procedures:

- The participants worked in seven groups. Each group discussed what questions it was going to ask parents, friends, idols, scientists, the government and why they wanted to ask these questions.

- Each group decided what questions it would present at the competition, wrote the questions in the given table and then memorized them.

- Group leaders were selected to ensure that all participants would have an opportunity. The group leaders took turns to report from memory and to justify the questions their group wanted to ask.

- The teacher jotted down each reported question and commented on it. Sometimes, he asked the group leader to provide more information.

- When a group leader in the corrective recast group reported a direct question (e.g. “Have you lied to me?”), the teacher first responded with an elicitation (“You want to ask them ...”). When the student said “I want to ask them ‘Have you’ / have they / whether have they lied to me”, the teacher repeated and then recast it. However, the teacher simply recast the student’s original utterance with no elicitation and no repetition in the implicit recast group.

- The rest of the class listened carefully to each proposed question and decided whether the question was a good one. They were encouraged to dispute any question they considered unnecessary.

- The teacher took this opportunity to recast any embedded question error.

- The teacher counted the votes for each question a group leader proposed.

- Finally, the teacher declared what the typical questions that they wanted to ask were.

Each task session lasted for 1.5 to 2 hours, during which the target structure errors were corrected using either explicit or implicit recasts in the experimental groups.

4.4.4 Recasts

This research discriminated between explicit and implicit recasts, and operationalized explicit recasts as corrective recasts.

Corrective recasts

Doughty and Varela (1998) defined corrective recasts as “(1) repetitions to draw attention followed by (2) recasts to provide the contrastive L2 forms” (pp. 123-124). When an
erroneous utterance was repeated, it was emphasized with added stress. When the correct form was provided in the recast, it was also emphasized. The added stress is “potentially a variable” and may “influence the extent to which the corrective force is explicit to learners” (Sheen, 2006, p. 364). Accordingly, corrective recasts in this research consisted of three elements: a repetition, a recast and salient supra-segmental features.

When a learner in the corrective recast group made a 3rd person -s or embedded question error, such as lack of the 3rd person -s token, over-use of the 3rd person -s token, inverted word order in embedded questions, and misuse of a question word, the teacher first repeated the erroneous part of the original utterance using a rising intonation, with a stress added to the error. When the learner failed to self-correct, the instructor then immediately reformulated the ungrammatical utterance with a falling intonation by correcting the error but keeping the intended meaning intact. The correct form was also emphasized. For instance, corrective recasts of a 3rd person -s error were as follows:

S: *The man wear a skirt.
T: * The man WEAR?
S: Wear a skirt.
T: The man WEARS.

The feedback on embedded question errors was performed in a slightly different way. When repeating and reformulating an ungrammatical utterance, the instructor kept addressing the learner as “you” rather than the original “I” or “we”. The exact repetition of “I” was likely to interrupt the communication flow and distract from the primary focus on meaning.

The second challenge was that the participants tended to use direct questions. When a participant produced only direct questions in a focused task, the teacher had to create opportunities for embedded questions through elicitations, such as “You want to know what…”, “You asked them…” and “You’d like people to tell you…”. An example of the corrective recasts of embedded question errors is as follows:

T: Okay. Now tell me what questions you will ask people about their eating habits.
S: “What do you usually have for breakfast?”
T: In the survey, you want to know what
S: * I want to know “What do you usually have for breakfast?”
T: * You want to know what DO YOU?
S: What do they.
T: Yes. You want to know what THEY usually HAVE.

According to Sheen’s (2006) taxonomy of recast characteristics, the corrective recasts on both structures were “declarative” in mode and focused on one particular structure in each instruction session. The corrective recasts involved several moves whereas the implicit recasts usually took the form of one move.

Implicit recasts

Implicit recasts are “the teacher’s reformulation of all or part of a student’s utterance that contains at least one error within the context of a communicative activity in the classroom” (Sheen, 2006, p. 365). They simply reformulate “a learner’s nontargetlike utterance into a targetlike one” keeping the intended meaning intact (Nassaji, 2009, p. 412). This research followed Lyster’s (2004) example, operationalizing implicit recasts as a turn-taking indicator and a reformulation. When a participant committed a target structure error, the instructor immediately took the next turn using words like “Oh” or “Ah” and then rephrased the entire ungrammatical utterance correcting the error. For example, the teacher recast a 3rd person -s error as follows:

S: *The man wear a skirt.
T: Ah, the man wears a skirt.

Implicit recasts of an embedded question error also involved the adjustment of personal pronouns. The teacher addressed the speaker as “you” when reformulating the erroneous utterance.

T: Okay. Now tell me what questions you will ask people about their eating habits.
S: I want to ask, “What do you usually have for breakfast?”
T: Oh. You want to ask people what they usually have for breakfast.

Unlike in Erlam and Loewen (2010), implicit recasts in this research employed a falling intonation because the implicitness of interrogative recasts is doubtful. Some researchers claim that interrogative recasts are implicit since they are usually perceived as confirmation checks rather than as corrective feedback (Lyster, 1998a) and they lead to less repair (Sheen, 2006) whereas others contend that the rising intonation renders the CF more salient as it leads the learner to check whether the teacher’s response is what he intended to say, resulting in a cognitive comparison (R. Ellis, 1995, cited in Erlam & Loewen, 2010, p. 879). In this study, both explicit and implicit recasts were declarative in mode.

4.4.5 Audio recorder and transcription
When a task was performed, the teacher wore a clip-on microphone and kept himself close to the task performer. The teacher-performer interaction was recorded on a voice recorder, which was set around a loudspeaker. The recorded CF episodes were transcribed using a broad transcription system but indicating extra stress, which is an essential characteristic of explicit recasts. Other discourse features indicated in the transcript are turns, intonation, speech overlap, pause and no response. Transcription conventions for these features are listed in Appendix C. Given below is an example of transcribed CF episodes.

S: She likes studying very much, so she seldom skip
T: she seldom SKIP?
S: -
T: She seldom skipS = any class.
S: = skips. En [yes].

4.4.6 Language tests
Acquisition was operationalized as the development of implicit and explicit knowledge of the target structures. Implicit knowledge is categorized as intuitive knowledge and it consists of formulaic knowledge and rule-based knowledge as R. Ellis (1994) stated. The former refers to “ready-made chunks of language” and the latter is “generalized and abstract structures
which have been internalized” (pp. 355-356). R. Ellis argued that this kind of knowledge is hidden and only becomes apparent when learners are using the language. Following Ellis et al. (2009), implicit knowledge was measured by means of an elicited imitation test (EI) and an oral production test (OP).

“Explicit L2 knowledge is that knowledge of rules and items that exist in an analyzed form so that learners are able to report what they know” (R. Ellis, 1994, p. 702). It is categorized as “declarative knowledge” or “metalingual knowledge”. In this research, explicit knowledge was measured by means of a written production test (WP) and an untimed grammaticality judgment test (UGJ) as these tests offered an opportunity for the learners to use explicit knowledge. These language tests were administered as pretest, immediate posttest (Posttest 1) and delayed posttest (Posttest 2) in the following order: EI, OP, WP and UGJ.

**Elicited imitation test**

In the EI test, the participants were required to listen to a statement, indicate on the answer sheet (see Appendix D) whether they agreed with it and finally repeat the statement. There were 18 belief statements for 3rd person -s (see Table 13) and 18 statements for embedded questions (see Table 14). Half of the statements were adapted from Ellis et al.’s (2009) replacing proper nouns and the other half were designed adhering to their principles. Half were grammatical and half were ungrammatical, but the participants were not informed that some of the sentences were ungrammatical. They were told that these sentences were all beliefs that they had to agree or disagree with.

Two thirds of the statements were for eliciting the target structure and one third served as distracters. The distracters covered regular past tense, possessive -s, plural -s, relative clauses, yes/no questions and question tags. These items were proofread by two native speakers of English and tried out on nine native speakers in Auckland prior to the experiment. The baseline data showed that eight of them reproduced all exemplars of the target structure and only one of them occasionally avoided using the target structure because he preferred to use plural subjects in reproducing the sentences. In general, these statements successfully created obligatory contexts for the target structure.

The stimuli for either target structure were regrouped according to their themes (i.e. a good teacher, a child, a wise man, etc.). This regrouping aimed to afford not only a focus on meaning but also opportunities to use 3rd person singular pronoun before the “-s” form. All
the stimuli were audio-recorded with a 12-second pause between items.

*Materials:*

“First of all, please report your name and group number to the voice recorder. Your name is… and your group number is…. Now, keep listening to the recording. You will hear altogether 18 statements, which you must agree with, disagree with or indicate you are not sure about. There is a 12-second pause between items. So, first decide whether you agree or not by marking A (true), B (not true) or C (not sure) on the answer sheet, and then repeat the sentence in CORRECT English. You have to repeat the speaker’s statement even if you don’t agree with it. Your repetition will be audio-recorded. Let’s start with eight practice items.”

*Table 12: EI Items for Practice*

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>*Every child needs good father.</td>
</tr>
<tr>
<td>2.</td>
<td>Children play basketball well and soccer badly in China.</td>
</tr>
<tr>
<td>3.</td>
<td>*Not everyone can to learn a second language.</td>
</tr>
<tr>
<td>4.</td>
<td>*It is more harder to learn Japanese than to learn English.</td>
</tr>
<tr>
<td>5.</td>
<td>We want to keep our country clean and green.</td>
</tr>
<tr>
<td>6.</td>
<td>China is greener and more beautiful than other countries.</td>
</tr>
<tr>
<td>7.</td>
<td>*People have been using natural gas since many years.</td>
</tr>
<tr>
<td>8.</td>
<td>People should report a car accident to the police.</td>
</tr>
</tbody>
</table>

*Note.* 1. These sentences were adapted from Ellis, Loewen, Elder, Erlam, Philp and Reinders’ (2009). 2. An asterisk * indicates an ungrammatical sentence.

“That’s all for our practice. Now the test begins.”
Table 13: EI Stimuli for Study 1

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>*A good teacher <strong>know</strong> every student <strong>grows</strong> at his own pace.</td>
</tr>
<tr>
<td>2.</td>
<td>He <strong>makes</strong> lessons interesting and <strong>cares</strong> about his students.</td>
</tr>
<tr>
<td>3.</td>
<td>*When he <strong>call</strong> a student by his full name, it always <strong>mean</strong> trouble.</td>
</tr>
<tr>
<td>4.</td>
<td>He <strong>loves</strong> comic books and is always reading them.</td>
</tr>
<tr>
<td>5.</td>
<td>*The Americans were first to land on the moon, <strong>didn’t they</strong>?</td>
</tr>
<tr>
<td>6.</td>
<td>*They bring more <strong>problem</strong> than <strong>solutions</strong> to the world.</td>
</tr>
<tr>
<td>7.</td>
<td>An American invented the software <strong>that changed the world</strong>.</td>
</tr>
<tr>
<td>8.</td>
<td>A child <strong>needs</strong> a role model, but not a supermodel.</td>
</tr>
<tr>
<td>9.</td>
<td>In China a child often <strong>visits</strong> clubs and <strong>drinks</strong> a lot.</td>
</tr>
<tr>
<td>10.</td>
<td>*A child <strong>take</strong> advice from everyone but his parents.</td>
</tr>
<tr>
<td>11.</td>
<td>*Nobody <strong>like</strong> <strong>Aftershock</strong>, the film by Feng Xiaogang.</td>
</tr>
<tr>
<td>12.</td>
<td>It <strong>tells</strong> the story of the Tangshan Earthquake.</td>
</tr>
<tr>
<td>13.</td>
<td>*The quake <strong>destroyed</strong> many houses and <strong>kill</strong> many people.</td>
</tr>
<tr>
<td>14.</td>
<td>Last <strong>year</strong>’s earthquakes shocked the whole world.</td>
</tr>
<tr>
<td>15.</td>
<td>A wise man <strong>keeps</strong> his eyes half-closed after marriage.</td>
</tr>
<tr>
<td>16.</td>
<td>*He <strong>return</strong> home as early as possible every day.</td>
</tr>
<tr>
<td>17.</td>
<td>*A responsible man <strong>take</strong> good care of his children.</td>
</tr>
<tr>
<td>18.</td>
<td>He used to ask, <strong>“Does the tie go with my shirt?”</strong></td>
</tr>
</tbody>
</table>

Note. An asterisk * indicates an ungrammatical sentence.
### Table 14: EI Stimuli for Study 2

1. Good students know **in what way they can learn English well.**
2. *They know **what do they want from school.**
3. *Students work hard in this school, aren’t they?
4. They are now planning **what jobs they will have in the future.**
5. It is a silly question to ask, “**Does a woman need to marry?”**
6. *Young women like **cigarettes and fast car.**
7. *They often ask **do their boyfriends really love them.**
8. *Wives always want to know **what are their husbands doing.**
9. *Aftershock is Feng Xiaogang’s latest film, produced in 2010.
10. People asked Feng Xiaogang **whether his film upset the victims.**
12. Before the quake people did not realize **what would happen.**
13. *We don’t care **what is President Obama like.**
14. We didn’t know **who Obama was** before his election.
15. *Obama explained **why didn’t he salute the U. S. flag that day.**
16. *Obama is not sure **whether will he be re-elected.**
17. My father lost the money **that he borrowed from the bank.**
18. He is not sure **whether he can earn money in a casino.**

*Note. An asterisk * indicates an ungrammatical sentence.*

---

**Procedures:**

- The students took the test one by one.
- The teacher explained the instructions in Chinese. The participant was told to first decide whether he/she agreed with the statement, making choices on the answer sheet and then to repeat it using CORRECT English.
- The teacher played the stimulus recording and switched on the voice-recording device.
- The participant practiced with eight distracter statements (see Table 12). The teacher made sure the participant indicated agreement/disagreement/not sure before repeating each item.
- The teacher administered the test, monitoring that the student was following the procedure and that the recorded stimuli were not replayed.
Oral production test

The other measure of implicit knowledge was an oral production test, in which participants orally reconstructed a passage they read. The reconstruction afforded an opportunity for “free production”. That is why Ellis et al. regard it as “the ideal measure of implicit knowledge” (2009, p.28).

The OP test for Study 1 employed a passage which was seeded with 26 exemplars of 3rd person -s and the test for Study 2 incorporated 12 exemplars of embedded questions. These two passages were proofread by two and piloted on nine native speakers of English. The native speakers reproduced 17 exemplars of 3rd person -s and eight exemplars of embedded questions on average.

Materials:

“Read the passage ‘A Typical Day in President Obama’s Life’ / ‘The Missing CD’ and try to understand it completely. If you come across any unfamiliar words, you can ask the teacher for help. After you finish reading, retell the passage to the voice recorder. You don’t have to use the original words, but you should include all the main events in the passage. Complete your retelling within FOUR minutes.”

Table 15: OP Passage for Study 1

<table>
<thead>
<tr>
<th>A Typical Day in President Obama’s Life</th>
</tr>
</thead>
<tbody>
<tr>
<td>President Obama usually gets up early in the morning. He goes for a workout at 6:45 a.m. He likes to do this before doing anything else because that way he can be sure that he has the time for a workout every day. After that, he eats breakfast with his family and then helps get his children ready for school before walking downstairs to work.</td>
</tr>
<tr>
<td>President Obama usually shows up at the Oval Office shortly before nine o’clock in the morning. On a typical day, he meets with his advisors and foreign leaders, signs bills and executive orders, works on the federal budget, and makes public appearances. He works until about 6:30 in the evening.</td>
</tr>
<tr>
<td>President Obama spends every evening with his family. He eats dinner with his family, hangs out with the kids and puts them to bed at about 8:30 p.m. And then he returns to work.</td>
</tr>
<tr>
<td>President Obama reads more briefing papers, does a lot of paperwork, and sometimes writes until nearly midnight. He knows that he has more mental energy at night. While much of the country is sleeping, the president thinks more clearly and makes his best decisions. He also dedicates some of his time to getting ahead — learning new things and improving himself. He expands his mind every night by reading, not by watching TV. It is interesting that he does not read about politics, foreign affairs or anything related to his “job”. He reads for pleasure. For example, he is currently reading Netherland, a novel about cricket.</td>
</tr>
</tbody>
</table>
Table 16: OP Passage for Study 2

The Missing CD

My sister Carly was standing in the hallway at school. She was looking through her purse for something. She kept asking herself where it was. I had no idea what she was looking for. She told me it was her CD. Without the CD, she couldn’t practice for the school play she was in. If she couldn’t practice, she would be kicked out of the play. I asked my sister where she might have left her CD. She was sure it was the gym. But when I suggested going there, she said she didn’t want me to go with her. I asked her whether she really wanted to find her CD. She groaned and said I could come along.

When we got to the gym, we saw a blonde girl. I asked who the girl was. The girl’s name was Shannon and she was unzipping her purse. It was identical to my sister’s. My sister immediately demanded to know why Shannon had taken her purse. Shannon said she did not understand what my sister was talking about. Then she accidentally dropped her purse and a CD fell out! Shannon looked surprised and said she did not know how the CD had got in there.

I asked Shannon to give me her purse. She asked me why I wanted it. I explained I wanted to check something. I looked inside and saw “Carly” written on a name tag. Then I looked inside the purse Carly was holding. My sister looked angrily at me and asked what I was doing. I looked inside Carly’s purse and saw “Shannon” on the name tag. Somehow the two girls had switched bags.

Shannon wanted to know where they could have switched the bags. It turned out that both Shannon and Carly had been having a meal in the school cafeteria and had put their bags on the same table. The two girls asked what they could do to thank me for sorting out their problem.

Note. When the texts were used in the experiment, translation of key words that were potentially new to the participants was provided (see Appendix E).

Procedures:

• The students took the test individually.

• The participant was told to read and then orally reproduce the passage.

• There was no time pressure for reading. However, when a participant was seen to be memorizing the passage, a time limit was set - 15 minutes in the pretest preparation, eight minutes in the immediate posttest and five minutes in the delayed posttest.

• When the participant came across unfamiliar words, he/she asked the teacher for help.

• The student was allowed to make notes of the main points but the notes were collected prior to his/her retelling. He/she was told not to write out full sentences.

• When the student was ready for retelling or preparation time was up, the teacher took the text and notes away.

• The student reported his/her name and group number and was then asked to reproduce the passage within four minutes.

• The student’s oral reproduction was recorded.
• When the student was stuck, the teacher reminded him/her of key words.

• The teacher encouraged the student to complete his/her reproduction of the passage even if time was up. The whole reproduction was analyzed later.

Written production test

The written production test immediately followed the oral production test. In this test, the participants were required to write out the passage he/she had read in the oral test. To reduce spelling difficulty, potentially unfamiliar words were supplied on the answer sheet (see Tables 17 and 18).

Materials:

Table 17: WP Test for Study 1

<table>
<thead>
<tr>
<th>The Written Production Pre/Post/Delayed Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>Name: _______________ Group: _______________</td>
</tr>
</tbody>
</table>

Instructions: Now you are required to write down what you have orally reproduced about the passage “A Typical Day in President Obama’s Life”. You can use your own words to write but please include all the main events in the passage you read. Since there is no time limit, you can take your time to make sure that you are writing in CORRECT English. Please double-check your writing before submitting it.

Here are some words that you might want to use in your writing.

- workout 健身
- show up 露面
- Oval Office 总统办公室
- executive 行政的
- federal budget 联邦预算
- hang out 闲逛
- briefing papers 简报
- dedicate 致力于
- get ahead 进步
- expand 拓展

Netherland 小说《荷兰》

cricket 板球
Table 18: WP Test for Study 2

The Written Production Pre/Post/Delayed Posttest

<table>
<thead>
<tr>
<th>Name:</th>
<th>Group:</th>
</tr>
</thead>
</table>

**Instructions:** Now you are required to write down what you have orally reproduced about the story “The Missing CD”. You can use your own words to write but please include all the main events of the story. Since there is no time limit, you can take your time to make sure that you are writing in **CORRECT** English. Please double-check your writing before submitting it.

Here are some words that you might want to use in your writing.

- Carly 我的姐姐
- hallway 门厅
- purse 小提包
- gym 健身房
- groan 抱怨
- blonde 金发女郎
- Shannon 姐姐的同学
- unzip 拉开
- identical 相同的
- accidentally 无意中
- name tag 名字标签
- switch 交换
- cafeteria 自助餐厅

**Procedures:**

- On completion of the oral production test, the participants were instructed to reproduce the passage they read in writing.

- The teacher reminded them that there was no time limit and they could take as much time as they needed to write out the passage in **CORRECT** English.

- When reconstructing the passage, the participants were not allowed to look at the text or any notes.

- They were encouraged to make corrections if they would like to.

- When a participant finished his/her writing, the teacher advised him/her to double check the writing before submitting it.

**Untimed grammaticality judgment test**

The untimed grammaticality Judgment was a “pen and paper” test. The test for each structure was comprised of 18 sentences (see Appendix F) and each sentence was presented on a separate page in a booklet. The participants were required to judge whether each sentence is grammatical or not, indicate the certainty of their judgment and the criteria for their judgment, and finally correct the structure if it is ungrammatical. Correction of ungrammatical items is likely to push them to make a “conscious analysis” (R. Ellis, 1991, cited in Loewen, 2009, p. 95) drawing on their explicit knowledge.
All the sentences were adapted from grammar sections of China’s National College Entrance Examinations from 2001 to 2008. The distracters included yes/no questions, tag questions, plural -s, relative clauses, comparative degree, since/for, possessive -s, regular past, indefinite articles, modal verbs, verb complements and adverb placement, which Ellis et al. (2009) categorized as difficult structures. Each UGJ test consisted of 12 sentences for the target structure and six sentences for distracters. Half of the test items were grammatical and the other half were ungrammatical. The UGJ tests were proofread and piloted on native speakers of English before implementation. All these native speakers agreed on the grammaticality/ungrammaticality of each sentence and the nature of the errors.

Presented below is a sample of the untimed grammaticality judgment test. For the complete UGJ tests, please refer to Appendix F.

**Materials:**
Table 19: Typical Items in UGJ Test

<table>
<thead>
<tr>
<th>Name:</th>
<th>Group:</th>
</tr>
</thead>
</table>

**Instructions:** In this section, you are expected first to decide whether each sentence is grammatically correct or incorrect for written English. Then, indicate the certainty of your judgment by marking a level in a 6-point scale. 0 means that you are totally uncertain and 5 means that you are 100% confident. After that, indicate whether you have judged it by “rule” or by “feel”. Finally, if the sentence is ungrammatical, please correct it. There is no time limit in this section, so you can take time to double-check your answers before turning over to the next page for a new item.

**[Typical items for Study 1]**

*My brother is fortunate to find a job that he love very much.*
- Grammaticality: A. correct  B. incorrect
- Certainty of judgment: very uncertain  0  1  2  3  4  5  very certain
- Judgment by:  A. rule    B. feel
- If ungrammatical, correct it here: ____________________.

The lawyer always wears a suit in spring.
- Grammaticality: A. correct  B. incorrect
- Certainty of judgment: very uncertain  0  1  2  3  4  5  very certain
- Judgment by:  A. rule    B. feel
- If ungrammatical, correct it here: ____________________.

**[Typical items for Study 2]**

*I want to ask the players why did they fail to win the game.*
- Grammaticality: A. correct  B. incorrect
- Certainty of judgment: very uncertain  0  1  2  3  4  5  very certain
- Judgment by:  A. rule    B. feel
- If ungrammatical, correct it here: ____________________.

*I doubt whether anyone in my class has a higher IQ than Jane.*
- Grammaticality: A. correct  B. incorrect
- Certainty of judgment: very uncertain  0  1  2  3  4  5  very certain
- Judgment by:  A. rule    B. feel
- If ungrammatical, correct it here: ____________________.

Note. An asterisk * indicates an ungrammatical sentence.

**Procedures:**

- The teacher explained the instructions in Chinese.
- The students read each sentence to judge whether it is grammatical or not, indicated the certainty of their judgment and the criteria for their judgment, and corrected the sentence if ungrammatical.
- The teacher reminded them that there was no time limit for this test and they could take as much time as they needed to judge each sentence.
- The teacher made sure that they completed each item before turning over to the next page in the test booklet.
• They were allowed to correct an answer before moving to a new item. Once having turned over the page, however, they were forbidden to return.

• When a student finished the test, he/she submitted the completed booklet to the teacher.

4.4.7 Working memory tests
Working memory is a psychological mechanism for simultaneously processing and retaining information. In Baddeley’s (2000) non-unitary model, working memory consists of the central executive, the phonological loop, the visuo-spatial sketchpad and the multimodal episodic buffer. WM and its components are limited in capacity. There are two common approaches to measuring working memory capacity, namely phonological short-term memory (PSTM) and complex working memory (CWM). The PSTM test asks participants to listen to and repeat a list of unrelated items, such as nonwords, words or numbers. The CWM test usually takes the form of a reading/listening/speaking span test, which requires participants first to understand sentences and then recall the symbols presented together with the sentences. The working memory test battery in this research was comprised of three sections: nonword span, digit span and listening span tasks. They were all administered in Chinese.

Nonword span test
Forty-eight 1-syllable Chinese nonwords were borrowed from Lee’s (2008) study: mang1, ran1, jiong1, rong1, min1, fou1, mou1, ruan1, niao1, rui1, kao1, yue2, shuan2, chuo2, zang2, suan2, nie2, ka2, le2, zhun2, pou2, dong2, dian2, sai3, nen3, nie3, run3, miu3, lue3, te3, die3, xiong3, teng3, nue3, niang3, ken4, lia4, fo4, qiong4, neng4, gei4, de4, qun4, dei4, nin4, diu4, kei4 (the numbers indicate Chinese lexical tones). All these pinyin are pronounceable but represent no character in Chinese. They were randomized to form sequences of two to nine nonwords and audio-recorded at the rate of one nonword per second, and then presented for the participants to repeat.

Materials:
“Listen to the recording carefully. When you hear the signal “Okay”, repeat the sequence of ‘words’ preceding ‘Okay’ immediately no matter whether you can understand it or not. You don’t have to follow the exact order of the ‘words’, but try to recall as many ‘words’ as you
can. If you fail one sequence, carry on with the next one. Continue until the teacher says ‘stop’.”

“Now, let’s have a try.”

Table 20: Nonword Sequences for Practice

<table>
<thead>
<tr>
<th>No. of Items</th>
<th>Content of Items</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>diu4, min1</td>
</tr>
<tr>
<td>2.</td>
<td>shuan2, teng3</td>
</tr>
<tr>
<td>3.</td>
<td>ran1, suan2</td>
</tr>
<tr>
<td>4.</td>
<td>dian2, qun4, zang2</td>
</tr>
<tr>
<td>5.</td>
<td>te3, nin4, dong2</td>
</tr>
<tr>
<td>6.</td>
<td>mie3, pou2, neng4</td>
</tr>
</tbody>
</table>

Note. The numbers following pinyin represent Chinese lexical tones.

“That’s the end of practice. Now, the test begins. Please report your name and group number to the voice recorder. Your name is… and your group number is…”
### Table 21: Nonword Sequences for the Test

<table>
<thead>
<tr>
<th>No. of Items</th>
<th>Content of Items</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>mou1, shuan2</td>
</tr>
<tr>
<td>2.</td>
<td>te3, chuo2</td>
</tr>
<tr>
<td>3.</td>
<td>dei4, jiong1</td>
</tr>
<tr>
<td>4.</td>
<td>rui1, gei4, nen3</td>
</tr>
<tr>
<td>5.</td>
<td>jiong1, qun4, niao1</td>
</tr>
<tr>
<td>6.</td>
<td>diu4, zhun2, xiong3</td>
</tr>
<tr>
<td>7.</td>
<td>nie2, kei4, run3, zang2</td>
</tr>
<tr>
<td>8.</td>
<td>lia4, miu3, le2, neng4</td>
</tr>
<tr>
<td>9.</td>
<td>niao1, mie3, kao1, sai3</td>
</tr>
<tr>
<td>10.</td>
<td>die3, le2, dei4, ruan1, miu3</td>
</tr>
<tr>
<td>11.</td>
<td>gei4, fou1, ruan1, xiong3, kao1</td>
</tr>
<tr>
<td>12.</td>
<td>de4, nue3, ken4, min1, mang1</td>
</tr>
<tr>
<td>13.</td>
<td>teng3, lue3, fo4, dong2, fou1, diu4</td>
</tr>
<tr>
<td>14.</td>
<td>nen3, yue2, run3, fo4, lue3, pou2</td>
</tr>
<tr>
<td>15.</td>
<td>dian2, teng3, suan2, pou2, ran1, qiong4</td>
</tr>
<tr>
<td>16.</td>
<td>mie3, rong1, diu4, chuo2, ka2, dei4, niang3</td>
</tr>
<tr>
<td>17.</td>
<td>ran1, shuan2, qiong4, mang1, sai3, ken4, zhun2</td>
</tr>
<tr>
<td>18.</td>
<td>ruan1, lue3, mou1, zhun2, gei4, zang2, yue2</td>
</tr>
<tr>
<td>19.</td>
<td>nen3, nuo1, teng3, ka2, fo4, rui1, miu3, run3</td>
</tr>
<tr>
<td>20.</td>
<td>qun4, dian2, de4, rui1, dong2, niang3, nue3, nuo1</td>
</tr>
<tr>
<td>21.</td>
<td>kei4, ran1, fou1, neng4, die3, chuo2, lia4, nie2</td>
</tr>
<tr>
<td>22.</td>
<td>yue2, suan2, nie2, kei4, zang2, mie3, rong1, nin4, ka2</td>
</tr>
<tr>
<td>23.</td>
<td>ken4, kao1, nin4, nue3, rong1, le2, te3, qiong4, niao1</td>
</tr>
<tr>
<td>24.</td>
<td>shuan2, pou2, jiong1, suan2, sai3, qun4, de4, min1, mang1</td>
</tr>
</tbody>
</table>

Note. The numbers following pinyin represent Chinese lexical tones.

**Procedures:**

- The participants took the test individually.
- Before the test, each participant listened to the Chinese instructions (see Appendix G) and then practiced with sequences of two nonwords and sequences of three nonwords (see Table 20 above).
- The nonword span test began with two-nonword sequences. For each sequence length, a participant had three trials.
• On hearing the signal “Okay”, the participant was asked to immediately repeat the sequence of nonwords he/she had just heard.

• When a participant started repeating before hearing “okay”, the teacher stopped him/her and reminded that he/she should wait for the signal word “Okay”.

• When a participant failed a sequence, the teacher encouraged him/her to try the next. After he/she failed all the three trials of a given sequence length, the teacher terminated the test.

Digit span test

Likewise, a digit span test required the participants to repeat digit strings of various lengths, from two numbers to nine numbers (see Table 22). These digit strings were randomly generated out of the numbers from 11 to 99 and were audio-recorded at the rate of one number per second. When recorded, they were read in Chinese pinyin. For example, the digit string “37 19 84” was read as “san1shi2qi1 yi1shi2jiu3 ba1shi2si4” (the numbers indicate Chinese lexical tones). Each string ended with “Okay”, a signal for recall.

Materials:

“First of all, please report your name and group number to the voice recorder. Your name is… and your group number is…. Then, listen to the recording carefully. When you hear the signal ‘Okay’, repeat the preceding string of numbers immediately. You don’t have to follow the exact order of the numbers, but try to recall as many numbers as you can. If you fail one sequence, carry on with the next one. Continue until the teacher says ‘stop’. Now the test begins.”
Table 22: Strings of Random Numbers for the Test

<table>
<thead>
<tr>
<th>No. of Items</th>
<th>Content of Items</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>24  93</td>
</tr>
<tr>
<td>2.</td>
<td>16  39</td>
</tr>
<tr>
<td>3.</td>
<td>51  32</td>
</tr>
<tr>
<td>4.</td>
<td>69  36  75</td>
</tr>
<tr>
<td>5.</td>
<td>21  44  91</td>
</tr>
<tr>
<td>6.</td>
<td>37  19  84</td>
</tr>
<tr>
<td>7.</td>
<td>25  82  77  56</td>
</tr>
<tr>
<td>8.</td>
<td>92  89  48  60</td>
</tr>
<tr>
<td>9.</td>
<td>20  59  93  62</td>
</tr>
<tr>
<td>10.</td>
<td>63  22  71  69  18</td>
</tr>
<tr>
<td>11.</td>
<td>72  35  57  83  90</td>
</tr>
<tr>
<td>12.</td>
<td>96  36  39  75  21</td>
</tr>
<tr>
<td>13.</td>
<td>41  33  68  19  74  32</td>
</tr>
<tr>
<td>14.</td>
<td>94  80  34  47  20  78</td>
</tr>
<tr>
<td>15.</td>
<td>12  24  21  62  33  83</td>
</tr>
<tr>
<td>16.</td>
<td>42  29  48  52  67  71  84</td>
</tr>
<tr>
<td>17.</td>
<td>82  77  85  14  73  96  51</td>
</tr>
<tr>
<td>18.</td>
<td>49  18  26  68  47  55  40</td>
</tr>
<tr>
<td>19.</td>
<td>13  96  28  36  62  75  79  26</td>
</tr>
<tr>
<td>20.</td>
<td>98  23  93  45  38  43  74  29</td>
</tr>
<tr>
<td>21.</td>
<td>54  27  57  16  22  83  95  87</td>
</tr>
<tr>
<td>22.</td>
<td>12  40  79  26  42  65  70  56  82</td>
</tr>
<tr>
<td>23.</td>
<td>86  30  76  50  11  31  87  53  99</td>
</tr>
<tr>
<td>24.</td>
<td>80  48  56  58  69  81  97  94  15</td>
</tr>
</tbody>
</table>

Procedures:

- The procedure for the digit span test was the same as that for the nonword span test.
- The Chinese instructions for this test can be found in Appendix H.

Listening span test
The listening span test required the participants to listen to sets of three to six sentences, judge the plausibility of each sentence after listening, and recall the last word of the sentences at the completion of each set. This test was administered in Chinese.

The stimulus sentences were taken from the Lewandowsky, Oberauer, Yang, and Ecker (2010) working memory test battery and then were revised as their sentences were designed for Taiwanese. Altogether 96 revised sentences were piloted on ten native speakers of Chinese in mainland China. Only those sentences whose plausibility/implausibility 90% of the Chinese speakers agreed on were selected. Finally, 36 plausible and 36 implausible sentences were accepted as stimuli for the test and eight sentences as stimuli for practice. The test stimuli were arranged randomly into 16 sets, which ranged from three to six sentences. These sets were finally randomized in order (see Appendix I).

As Waters and Caplan (1996) noted, participants’ WM scores should be based not only on the number of words recalled but also on their reaction time and accuracy of acceptability judgment if storage and processing performances are both to be taken into consideration. To capture reaction time, the test was computerized. The stimulus sentences were audio-recorded and presented using the DMDX experiment software. When the participants heard a sentence, they judged its plausibility and pressed the button marked GOOD or BAD. When they finished a set, ranging between three and six sentences, the word “RECALL” appeared on the screen. Upon seeing the signal, they started to write out sentence-final words in the whole set on the answer sheet (see Appendix J).

Materials:

“In this test, you will hear Chinese sentences in groups. Each group contains three, four, five or six sentences. Hearing each sentence, you are expected to do two things. Firstly, decide whether the sentence makes good sense or not. If you think it is a good sentence, press the button labeled GOOD. If you don’t think it is acceptable, press the button labeled BAD. Secondly, please remember the last word of each sentence you heard. When you have completed all the sentences in a group, the word ‘RECALL’ will appear on the screen. Upon seeing ‘RECALL’, please write down the last word of each sentence you heard on the answer sheet. The sequence in which you write the words is not important. Note that in this test, reaction time, plausibility judgment, and recall accuracy are equally important.”

“Now you have an opportunity to practice with some sentences.”
Table 23: A Set of Sentences for Listening Span Practice

<table>
<thead>
<tr>
<th>No. of Sets</th>
<th>Content of Sets</th>
</tr>
</thead>
<tbody>
<tr>
<td>练习二</td>
<td>*他从小就注重锻炼，所以就弄坏了身体。 (He takes regular exercise from his childhood, <strong>and so</strong> the exercise ruins his health.)</td>
</tr>
<tr>
<td>(Practice Set 2)</td>
<td>李先生工作很辛苦，但收入却不太理想。 (Mr. Li works hard, but his income is unsatisfactory.)</td>
</tr>
<tr>
<td></td>
<td>*过去，北京的公交车站牌将要加注英文。 (In the past, bus station signs in Beijing <strong>will</strong> be appended with English notes.)</td>
</tr>
</tbody>
</table>

Note. 1. Here listed is only one set of stimuli sentences. Other sets for practice can be found in Appendix I.  
2. An asterisk * indicates an implausible sentence.

“That’s the end of our practice. Now, the listening span test begins.”

Table 24: A Set of Sentences for Listening Span Test

<table>
<thead>
<tr>
<th>No. of Sets</th>
<th>Content of Sets</th>
</tr>
</thead>
<tbody>
<tr>
<td>测试一</td>
<td>我们待人应心胸宽广，待己要严格要求。 (Towards others, we should be lenient; with ourselves, we should be strict.)</td>
</tr>
<tr>
<td>(Set 1)</td>
<td>*经过漫长的急救，小狗总算救活了兽医。 (After a long period of first aid, the puppy finally saved the vet’s life.)</td>
</tr>
<tr>
<td></td>
<td>*我国天然货物很多，种类只有煤和石油。 (China has a wealth of natural commodities, <strong>and there are only two kinds</strong>: coal and oil.)</td>
</tr>
</tbody>
</table>

Note. 1. Here listed is only one set of stimuli sentences. Other sets for the test can be found in Appendix I.  
2. An asterisk * indicates an implausible sentence.

Procedures:

- The participants were tested individually.
- Before the test, each participant listened to the Chinese instructions and then practiced with three sets of sentences.
- The teacher made sure that each participant was familiar with the procedure before the test began.
- The participant completed 16 sets of sentences during the test. For each set, he/she listened to each sentence, judged its acceptability, remembered its final word, and
when seeing “RECALL”, immediately wrote out the sentence-final words.

- When a participant started to write the sentence-final words before seeing the signal “RECALL”, the teacher explained that he/she should do it only after the signal appeared.

- When the participant failed a set, the teacher encouraged him/her to carry on with the next.

- On completion of the test, the teacher made sure that all the data, electronic and written, were saved under the participant’s name.

4.4.8 Exit questionnaire

The exit questionnaire was delivered via computers. It sought for information about what role the participants believed a teacher should play in task-based lessons, what they thought the teacher was doing in class, whether they noticed the teacher’s corrections, what attitudes they had towards error correction, and whether their affective disposition, such as their belief about a language teacher’s role and their attitudes towards error correction, could account for their noticing of the focus of feedback. There were altogether six questions in the questionnaire and each question was presented on a single webpage. Only when a participant had submitted the answer to a question, could he/she access the next one. The participant could not return to the previous question. This survey was conducted in Chinese (see Appendix K). Its English version is as below.

Materials:
### Table 25: Exit Questionnaire (English Version)

**Exit Questionnaire**

Thank you for participating in this research in the last three months and providing valuable information through this questionnaire. Your questionnaire will NOT be shown to your university authority or any other third party and the information you provide will be used anonymously for academic purposes. Please be as candid as possible when answering this questionnaire.

There are altogether six questions. Only when you have submitted the answer to the current question, can you access the next one. Please check your answer before submitting it because you cannot return to correct it.

1. What is your name and which group are you in?

2. What did you think the main purpose of the lessons was?

3. What did the teacher do in the lessons?

4. When you were performing communicative tasks, did you notice the teacher correcting your / a student’s grammatical error?

   If yes, can you recall any of the specific errors the teacher corrected?

   If no, go directly to Question 5.

5. Do you think the teacher’s correction is helpful? Why or why not?

6. Do you have other comments / suggestions on the whole project?

---

**Procedures:**

- Two Information Technology students developed a program and tested it for the exit questionnaire.
- The participants read the Chinese instructions and then answered the questionnaire on computers.
- Only after submitting the answer to a question, could the participants access the next one.
- They could not return to correct their previous answers.
When a participant completed his/her questionnaire, the teacher checked whether the participant’s electronic answers had been saved properly.

4.5 Data analysis

4.5.1 Analysis of recasts and uptake

Recording of the task-based lessons was examined to identify CF episodes using Lyster and Ranta’s (1997) framework. These CF episodes were transcribed and then coded with the following categories: “corrective recasts”, “implicit recasts”, “no uptake”, “uptake-without-repair” and “uptake-with-repair”. “Repair” was sub-coded in terms of “repetition” and “incorporation”. Corrective recasts and implicit recasts composed the treatment as defined in Section 4.4.4. The different forms of uptake, defined 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), were as follows:

1) Uptake-without-repair: “uptake that results in an utterance that still needs repair”.

2) Uptake-with-repair: “correct reformulation of an error as uttered in a single student turn”.
   a) Repetition: “a student’s repetition of the teacher’s feedback when the latter includes the correct form”.
   b) Incorporation: “a student’s repetition of the correct form provided by the teacher, which is then incorporated into a longer utterance produced by the student”.

These codes were quantified in terms of frequency. A repair/incorporation rate for each feedback receiver was calculated by dividing the number of repairs/incorporations by the total of received corrective/implicit recasts. A second rater coded and scored the IR group’s uptake-with-repair data for 3rd person -s, and the inter-rater reliability reached $r = .98$.

4.5.2 Analysis of implicit knowledge

Oral data from the elicited imitation and oral production tests were transcribed using a broad transcription method. Pauses were marked as they could indicate an attempt on the part of a participant to use their explicit knowledge.
The transcripts were first coded with the following categories: “obligatory context”, “correct suppliance”, and “suppliance in a non-obligatory context”. It should be noted that when a participant produced only words or nothing in his imitation of an EI item, this was still regarded as an “obligatory context”. When a participant produced a correct form of the target structure after multiple attempts, this utterance was not coded as “correct suppliance” because it did not reflect implicit knowledge for which the EI and OP tests were designed. Some participants failed to complete the retelling within four minutes but were allowed to finish. The whole reproduction of each participant was coded and scored but only if the target structure was produced without a pause. The detailed coding system for these oral tests is shown in Appendix L and Appendix M.

Then, a target-like use analysis was conducted using the formula “\([\text{Number of correct suppliance in contexts} / (\text{Number of obligatory contexts} + \text{Number of suppliance in non-obligatory contexts})] * 100 = \text{per cent accuracy}\)” (Pica, 1983, p.71). A second rater coded and scored 5% of the EI data for 3rd person -s and the OP data for embedded questions, and the inter-rater reliability was \(r = .87\) for the former and \(.91\) for the latter.

4.5.3 Analysis of explicit knowledge

Scoring written reproduction

The written reproduction was coded and analysed in the same way as the oral reproduction. The only difference was that any correct use of the target structure, even one that involved a correction (i.e. an exemplar that was once erroneous but was corrected to be grammatical before submission), was categorized as “correct suppliance” because the written production test was designed for explicit knowledge. A second rater coded and scored 6% of the WP data for 3rd person -s, and the inter-rater reliability was \(r = .98\).

Scoring the untimed grammaticality judgment

The untimed grammaticality judgment responses were analysed taking the participants’ correction of the sentences into consideration. When a participant marked a grammatical sentence as “incorrect” and attempted to correct a feature other than the target structure, he/she in fact had accepted the targeted part of the sentence, so he/she scored a point for judgment. When a participant judged an ungrammatical sentence as “incorrect” but mislocated the error, he/she had actually accepted the ungrammatical exemplar of the target structure as “correct”, so he/she was not given a point for this item. When a participant
located the error correctly producing no correct form, he/she scored a point for judgment but was not awarded 0.5 point for the correction. The scoring scheme for the UGJ data is appended in Appendix N.

A second rater marked 6% of the UGJ data for embedded questions, and the inter-rater reliability was $r = .97$. Each participant’s scores were finally converted into percentages. Their scores for grammatical and ungrammatical sentences were also calculated separately.

### 4.5.4 Factor analysis of implicit and explicit knowledge scores

A principal-components factor analysis was conducted of the participants’ EI, OP, WP and UGJ pretest scores for 3rd person -s. Outliers in the pretest scores were first identified and winsorized by 10%, that is, set at the 10th percentile. Then, factorability was examined by determining that a correlation coefficient was above .3 ($r = .34$), sampling adequacy was above .6 ($KMO = .66$), there was significant sphericity ($\chi^2(6) = 66.67, p < .05$) and that all diagonals of the anti-image correlation matrix were above .3.

Given the well-met conditions, a principal-components analysis was performed. Based on an eigen value of greater than 1.0, only one component was extracted. This factor, with an eigen value of 2.0, explained 50.1% of the variance. Communalities represent the proportion of the variance in the original variables that is accounted for by the factor solution, so the variable with a communality value of less than .5 should be removed from the component analysis. After the EI and OP pretest scores were removed, there were only WP and UGJ pretest scores left in the factor solution. The results indicate that WP and UGJ pretests can be represented by one factor, explicit knowledge.

Table 26: Factor Loadings and Communalities: A One Factor Solution ($N = 109$)

<table>
<thead>
<tr>
<th></th>
<th>Factor 1</th>
<th>Communality</th>
</tr>
</thead>
<tbody>
<tr>
<td>EI pretest</td>
<td>.61</td>
<td>.37</td>
</tr>
<tr>
<td>OP pretest</td>
<td>.68</td>
<td>.46</td>
</tr>
<tr>
<td>WP pretest</td>
<td>.78</td>
<td>.61</td>
</tr>
<tr>
<td>UGJ pretest (ungrammatical items)</td>
<td>.75</td>
<td>.56</td>
</tr>
</tbody>
</table>

Note. EI=elicited imitation; OP=oral production; WP=written production; UGJ=untimed grammaticality judgment.
When we asked for a solution with two factors, a principal-components factor analysis using varimax rotation provided the factor structure shown in Table 27. As this table shows, all the pretests had a communality value of above .5 in the two factor solution. Component 1 included the WP and UGJ tests and Component 2 included the EI and OP tests. In line with R. Ellis (2005b), Component 1 was labelled “explicit knowledge” and Component 2 “implicit knowledge”.

Table 27: Factor Loadings and Communalities: A Two Factor Solution

<table>
<thead>
<tr>
<th></th>
<th>Factor 1</th>
<th>Factor 2</th>
<th>Communality</th>
</tr>
</thead>
<tbody>
<tr>
<td>EI pretest</td>
<td>.86</td>
<td>.75</td>
<td></td>
</tr>
<tr>
<td>OP pretest</td>
<td>.74</td>
<td>.62</td>
<td></td>
</tr>
<tr>
<td>WP pretest</td>
<td>.85</td>
<td>.76</td>
<td></td>
</tr>
<tr>
<td>UGJ pretest (ungrammatical items)</td>
<td>.87</td>
<td>.77</td>
<td></td>
</tr>
</tbody>
</table>

Note. 1. EI=elicited imitation; OP=oral production; WP=written production; UGJ=untimed grammaticality judgment. 2. Factor loadings <.3 are suppressed.

4.5.5 Analysis of working memory

Phonological short-term memory

Only the nonword span data were scored because the digit span data were found to be invalid. Around 20% of the participants had “cheated” by jotting down the sequences of numbers as they heard them.

The participants had three trials at each sequence length. A trial was accepted as “correct” when a participant repeated back all the nonwords in the sequence irrespective of whether the repetition was in the original order. When the participant failed all the three trials of a certain sequence length, the test was terminated. Then, the previous sequence length was counted as the participant’s nonword capacity if all the trials were correct at that length. When any one of the trials was erroneous at the previous sequence length, his/her PSTM score was the previous length minus 0.3. If two trials of the previous length were erroneous, his/her score was the previous length minus 0.6.

Complex working memory
The participants’ listening span scores were based on their reaction times when they judged the plausibility of each sentence, the accuracy of their judgments and the accuracy of the recalled words. Each participant’s mean reaction time for correctly judged sentences was multiplied by -1 because a longer time indicated a slower response. The number of correctly judged sentences and the number of correctly recalled sentence-final words were counted for each participant. The scores from each measure were transformed into a z-score and then the 3 z-scores were averaged into a composite score.

4.5.6 Exit questionnaire

The exit questionnaire data were coded in terms of the following categories: “noticing of CF”, “noticing of 3rd person -s”, “noticing of embedded questions”, “support for CF” and “dislike of CF”. The frequency of each category was used to provide support in the discussion of the language test and uptake results. A Pearson chi-square ($\chi^2$) test of the reported noticing was conducted to examine whether corrective recasts and implicit recasts led to noticing to a different extent.

4.5.7 Statistical procedures for data analysis

Scores for uptake, implicit knowledge, explicit knowledge and working memory were entered into Statistical Package for the Social Sciences (SPSS) version 21. First of all, normality of the data was assessed graphically and numerically. When data points are all close to the diagonal line in the normal Q-Q plot, this suggests a normal distribution. Numerically, the absolute value of “skewedness / SE of skewedness” and the absolute value of “kurtosis / SE of kurtosis” are both less than 1.96 in a normal distribution. These criteria were used to test the normality of data distribution for the whole sample and for each group. When the scores were found to be normally distributed parametric tests were used in subsequent analyses, but when they were shown to be not normally distributed, non-parametric tests were applied.

**RQ1. What effects do corrective recasts and implicit recasts have on learners’ implicit knowledge of L2 grammatical features (3rd person -s and embedded questions)?**

A mixed 3 x 4 ANOVA of the participants’ implicit knowledge scores measured by EI and OP tests was carried out with “group” as a between-subject variable and “time” as a within-subject variable. When the group x time interaction was significant, the main effect of time
was examined in each group by means of separate repeated-measures ANOVAs / non-parametric Friedman tests (with skewed distribution). The main effect of group at each testing point was analyzed using one-way ANOVAs / non-parametric Kruskal-Wallis $H$ tests (with skewed distribution). When there was a significant between-group difference in the pretest, posttest scores were compared across groups by means of parametric/non-parametric ANCOVAs with pretest scores as the covariate. The detailed parametric or non-parametric statistical procedures used in each analysis are described in the Results and Discussion chapters. These inferential analyses help to address what effect corrective/implicit recasts have on the learners’ implicit knowledge and whether there is any difference between the effects of explicit and implicit recasts, that is, whether explicitness plays a role in the efficacy of recasts.

**RQ2.** What effects do corrective recasts and implicit recasts have on learners’ explicit knowledge of L2 grammatical features?

Similarly, a mixed 3 x 4 ANOVA of explicit knowledge scores measured by WP and UGJ (ungrammatical item) tests was first conducted in order to explore whether there was an interaction between feedback groups and testing times. After the group-time interaction was found significant, the main effect of time was examined using a parametric/non-parametric repeated-measures ANOVA, and the main effect of group employed a parametric/non-parametric ANOVA or ANCOVA. These analyses aimed to explore the role of recasts and CF explicitness in the development of explicit knowledge.

**RQ3.** Is there any difference in uptake-with-repair following corrective recasts and implicit recasts?

To answer the question whether repair frequency differed following explicit recasts and implicit recasts, the corrective recast (CR) and the implicit recast (IR) groups’ uptake-with-repair and non-repair (uptake-without-repair plus no uptake) turns were tallied and a Pearson chi-square ($\chi^2$) test of the frequencies was conducted. The numbers of participants who successfully repaired (repairers) and those who failed to repair their errors (non-repairers) in the two experimental groups were also compared via a chi-square test.
RQ4. Is there any relationship between learners’ uptake-with-repair and their subsequent acquisition of implicit and explicit knowledge?

For the CR group, bivariate correlations were conducted to see whether there was a relationship between repair rate and gain scores for implicit/explicit knowledge. Gain scores were the improvement from pretest to immediate/delayed posttest. For the IR group, however, linear correlation analyses were not appropriate because more than two thirds of the participants did not repair their errors and consequently the data had a nonlinear distribution. Therefore, the IR group’s repairers and non-repairers’ gain scores were compared by means of independent-samples t tests.

RQ5. What role does working memory play in corrective feedback involving recasts?

Whether differences in the learners’ working memory mediated their ability to repair was investigated separately for the CR and IR groups given the significant difference in repair rate between the two groups. Bivariate correlation analyses were conducted for the CR group in the case of 3rd person -s but independent-samples t tests of the repairers and non-repairers’ WM scores were carried out in the case of embedded questions as this group’s repair data were not linearly distributed. For the IR group, independent-samples t tests were computed in the case of 3rd person -s due to the nonlinear distribution of repair rates, but no analysis was conducted in the case of embedded questions because of the extremely small sample of repairers.

Whether differences in the learners’ working memory mediated their acquisition of target structures was examined using bivariate correlation analyses. Correlations of WM capacity and pretest-posttest gain scores were performed for the CR and IR groups separately. In both conditions, the correlations were run for two different target structures, 3rd person -s and embedded questions. For both structures, implicit and explicit knowledge scores were used.

RQ6. Is there any difference between the two target structures in uptake rate following explicit/implicit recasts, recasts-driven implicit/explicit knowledge development, and
the role of working memory in CF involving recasts?

No particular statistical procedure was performed for this research question, but all the above analyses were conducted separately for 3rd person -s and embedded questions. Differences and similarities between the two target structures in uptake rate, implicit and explicit knowledge development and function of working memory were observable from the results for the other research questions.

The statistical procedures for each research question are summarized in Table 28. To recap briefly, parametric/non-parametric means comparison procedures were carried out to examine the effect of recasts on language knowledge; chi-square tests were performed to investigate the influence of explicit/implicit recasts on immediate repair; and correlation analyses or t tests were conducted to study the relationship between WM, repair and acquisition.

Table 28: A Summary of Statistical Procedures for Data Analysis

<table>
<thead>
<tr>
<th>Research Questions</th>
<th>Procedures of Data Analysis</th>
<th>Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3 x 4 ANOVA</td>
<td>EI and OP</td>
</tr>
<tr>
<td></td>
<td>repeated-measures ANOVA / Friedman test</td>
<td></td>
</tr>
<tr>
<td></td>
<td>one-way ANOVA / Kruskal-Wallis H test</td>
<td></td>
</tr>
<tr>
<td></td>
<td>or ANCOVA</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>3 x 4 ANOVA</td>
<td>WP and UGJ (ungrammatical item)</td>
</tr>
<tr>
<td></td>
<td>repeated-measures ANOVA / Friedman test</td>
<td></td>
</tr>
<tr>
<td></td>
<td>one-way ANOVA / Kruskal-Wallis H test</td>
<td></td>
</tr>
<tr>
<td></td>
<td>or ANCOVA</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Pearson chi-square ($\chi^2$)</td>
<td>repair and non-repair frequency</td>
</tr>
<tr>
<td></td>
<td>repairer and non-repairer population</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>bivariate correlations (CR group)</td>
<td>repair rate and language gains</td>
</tr>
<tr>
<td></td>
<td>independent-samples t tests (IR group)</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>bivariate correlations for -s, independent-samples t test for questions (CR group)</td>
<td>WM scores and repair rate</td>
</tr>
<tr>
<td></td>
<td>independent-samples t test for -s (IR group)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>bivariate correlations (CR and IR groups)</td>
<td>WM scores and gain scores</td>
</tr>
<tr>
<td>6</td>
<td>review of the above procedures</td>
<td>all the above data</td>
</tr>
</tbody>
</table>
In addition, effect sizes (ES) were calculated. For ANOVA analyses, eta squared ($\eta^2$) was computed by dividing the effect of interest (SS\_Effect) by the total amount of variance (SS\_Total):

$$\eta^2 = \frac{SS\_Effect}{SS\_Total}.$$  

For ANCOVA analyses, partial eta squared ($\eta_p^2$) was computed by dividing the effect of interest by the sum of the variance attributable to the effect (SS\_Effect) and the error variability (SS\_Residual):

$$\eta_p^2 = \frac{SS\_Effect}{SS\_Effect + SS\_Residual}.$$

An $\eta^2$ value of .01 to .06 was interpreted as small ES, .06 to .14 as medium ES, and above .14 as large ES.

Cohen’s $d$ (1988) was calculated using the formula of “$d = (\text{Mean 1} - \text{Mean 2}) / \text{SD}$” corrected by pooled standard deviation for a between-group difference (Field, 2009), corrected by the correlation coefficient for a pretest-posttest difference (Morris & DeShon, 2002), and corrected by the correlation of covariate and dependant variable for a difference in the case of ANCOVA (Lipsey & Wilson, 2001). A $d$ value of .2 to .5 was considered as small ES, .5 to .8 as medium ES, and above .8 as large ES.

For chi-square tests, phi ($\phi$) coefficient was used as an effect size, which was the square root of the chi-squared statistic divided by the sample size: $\phi = \sqrt{\chi^2 / N}$ (Field, 2009). A $\phi$ value of .1 to .3 was regarded as small ES, .3 to .5 as medium ES, and above .5 as large ES.

For correlational analyses, the value of correlation coefficient $r$ functions as an effect size (Field, 2009). An $r$ value of .1 to .3 was regarded as small ES, .3 to .5 as medium ES, and above .5 as large ES. The details of how the effect sizes were calculated and interpreted are also given in each results chapter.
Chapter 5  Effect of Recasts on Implicit Knowledge: 
Results and Discussion

Chapter 5 explores the relationship between feedback types and implicit knowledge development.

RQ1. What effects do corrective recasts and implicit recasts have on learners’ implicit knowledge of L2 grammatical features (3rd person -s and embedded questions)?

1a. What effect do corrective recasts have on learners’ implicit knowledge of L2 grammatical features?

1b. What effect do implicit recasts have on learners’ implicit knowledge of L2 grammatical features?

1c. Is there any difference between the effects of corrective and implicit recasts on learners’ implicit knowledge?

5.1 Statistical treatment

To address the question whether or not recasts facilitate learners’ implicit knowledge development, the results of elicited imitation (EI) and oral production (OP) tests were entered into the SPSS package (Version 21). Both tests required participants to complete a task within the specified time span, and the time pressure may have pushed them to draw on implicit knowledge. This assumption is justified by previous pertinent research (e.g. R. Ellis, 2005; Ellis et al., 2009) and the results of a principal-components factor analysis of the EI, OP, written production (WP) and untimed grammaticality judgment (UGJ) pretest scores for 3rd person -s in this study. WP and UGJ ungrammatical item scores constituted Factor 1 – explicit knowledge – whereas EI and OP scores comprised Factor 2 – implicit knowledge.

A mixed 3 x 4 analysis of variance (ANOVA) of the EI/OP scores was conducted with “group” as a between-subject variable and “time” as a within-subject variable. When the interaction between time and group was significant, the main effect of time was examined in each group by means of separate repeated-measures ANOVAs or non-parametric Friedman tests, and the main effect of group at each testing point was analysed using one-way ANOVAs or non-parametric Kruskal-Wallis $H$ tests. When the analysis of pretest scores
reported a significant between-group difference, posttest scores were compared across groups by means of parametric or non-parametric ANCOVAs with the pretest scores as a covariate. Finally, post-hoc comparisons were carried out to investigate whether corrective and implicit recasts have differential effects on L2 learning.

For a significant time x group interaction, an effect size (partial eta squared $\eta^2_{p}$) is reported to reveal the magnitude of interaction. A $\eta^2_{p}$ value of .01 to .06 is evaluated as small ES, .06 to .14 is medium ES, and above .14 is large ES. A $\eta^2_{p}$ value is also reported for the significant main effect of time/group. For a significant difference between groups or a significant gain from pretest to posttest, Cohen’s $d$ is calculated using the formula of “$(\text{Mean 1} - \text{Mean 2}) / \text{SD}$” corrected by pooled standard deviation to examine the magnitude of a between-group difference (Field, 2009), corrected by the correlation coefficient to examine the magnitude of a pretest-posttest difference (Morris & DeShon, 2002), and corrected by the correlation of covariate and dependent variable to examine the magnitude of treatment in the case of ANCOVA (Lipsey & Wilson, 2001). A $d$ value of .2 to .5 is interpreted as small ES, .5 to .8 is medium ES, and any value above .8 is large ES.

Besides, a Pearson chi-square ($\chi^2$) test of reported noticing in the exit questionnaire was conducted to see whether corrective recasts and implicit recasts led to noticing to a different extent. The effect size (ES) for a chi-square test is reported using a phi ($\phi$) value, in which .1 to .3 is small ES, .3 to .5 is medium ES, and .5 to 1 is large ES.

### 5.2 Results

The results of descriptive and inferential statistics for 3rd person -s (the target of Study 1) and embedded questions (the target of Study 2) are reported respectively.

#### 5.2.1 Results for 3rd person -s: Elicited imitation

**Descriptive statistics**

Means and standard deviations (SD) of the participants’ scores in elicited imitation (EI) tests of 3rd person -s are summarized in Table 29. The corrective recast group had a mean score of 33.34 in the pretest, gained a higher mean of 49.35 in posttest 1 and then sustained its improvement in posttest 2 ($M = 50.75$) as illustrated in Figure 3. The implicit recast group’s scores also improved from pretest ($M = 39.26$) to posttest 1 ($M = 50.16$) and kept a sustained growth in posttest 2 ($M = 54.16$). The task control group scored the lowest with a mean of
30.49 in the pretest, but it followed the same development pattern (M = 42.28 in posttest 1 and M = 44.66 in posttest 2) as the experimental groups. In contrast, the test control group’s scores edged down from pretest (M = 43.35) to posttest 1 (M = 42.51) and further down in posttest 2 (M = 41.47).

Table 29: Descriptive Statistics of Scores for 3rd Person -s in EI Tests

<table>
<thead>
<tr>
<th>Group</th>
<th>Pretest</th>
<th>Posttest 1</th>
<th>Posttest 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
</tr>
<tr>
<td>CR group (n=32)</td>
<td>33.34</td>
<td>17.06</td>
<td>49.35</td>
</tr>
<tr>
<td>IR group (n=32)</td>
<td>39.26</td>
<td>15.45</td>
<td>50.16</td>
</tr>
<tr>
<td>TkC group (n=22)</td>
<td>30.49</td>
<td>18.04</td>
<td>42.28</td>
</tr>
<tr>
<td>TtC group (n=21)</td>
<td>43.35</td>
<td>17.70</td>
<td>42.51</td>
</tr>
<tr>
<td>Total (n=107)</td>
<td>36.49</td>
<td>17.34</td>
<td>46.80</td>
</tr>
</tbody>
</table>

Note. CR = corrective recast; IR = implicit recast; TkC = task control; TtC = test control.

Figure 3: Means for 3rd Person -s in EI Tests over Time by Group
Inferential statistics

Omnibus results

A mixed 3 x 4 ANOVA of the scores for accuracy of 3rd person -s in EI tests was conducted across groups over time, with group as a between-subject variable and time as a within-subject variable. The omnibus analysis showed that the interaction between time and group was significant, \( F(6, 206) = 3.86, p = .00, \eta^2_p = .10 \).

Results for the effect of time in each group

The significant time x group interaction necessitates a repeated-measures ANOVA for each group as a follow-up procedure. The ANOVA of the CR group’s EI scores found a significant difference between testing times, \( F(2, 62) = 25.85, p = .00, \eta^2_p = .46 \). Bonferroni post-hoc comparisons revealed a significant difference from pretest to posttest 1, \( p = .00, d = .96 \), and a significant difference between pretest and posttest 2, \( p = .00, d = 1.14 \), but no significant difference between posttests. For the IR group, a repeated-measures ANOVA was carried out with degrees of freedom corrected using Greenhouse-Geisser estimates of sphericity (\( \varepsilon = .81 \)) because the assumption of sphericity had been violated as determined by the Mauchly’s test, \( \chi^2 (2) = 8.19, p = .02 \). The analysis for the IR group also showed a significant difference over time, \( F(1.61, 50.04) = 15.64, p = .00, \eta^2_p = .34 \). Bonferroni post-hoc comparisons revealed a significant difference from pretest to posttest 1, \( p = .00, d = .92 \), and between pretest and posttest 2, \( p = .00, d = 1.02 \), but no significant difference between posttests. For the task control group, there was also a significant difference between testing points, \( F(2, 42) = 9.65, p = .00, \eta^2_p = .32 \). Post-hoc analyses showed a significant improvement from pretest to posttest 1, \( p = .02, d = .66 \), and a significant improvement between pretest and posttest 2, \( p = .00, d = .84 \), but no significant difference between posttests. For the test control group, no significant difference was found between testing points, \( F(2, 40) = .10, p > .05, \eta^2_p = .01 \).

Results for the effect of group at each testing point

A one-way ANOVA of EI pretest scores revealed a significant between-group difference in accuracy of 3rd person -s prior to the treatment, \( F(3, 103) = 2.72, p = .05, \eta^2 = .07 \). Bonferroni post-hoc comparisons found no significant between-group difference, but the task control group’s EI pretest scores were higher than the test control group’s, with a medium to large effect size, \( p = .09, d = .72 \). Given the close-to-large effect size, EI posttest scores were
compared across groups using ANCOVAs with pretest scores as a covariate.

A preliminary ANCOVA for EI posttest 1 controlling for the effect of the participants’ initial knowledge indicated that the assumption of homogeneity of variance was violated, $F(3, 103) = 3.88, p = .01$. Taking into consideration the fact that the ratio of the largest group variance to the smallest group variance was less than 3, the ANCOVA was still carried out for posttest 1. The results indicated a significant between-group difference among adjusted means (see Table 30) in posttest 1, $F(3, 102) = 3.16, p = .03, \eta^2_p = .09$. Pairwise comparisons determined that the CR group’s adjusted mean 51.13 was significantly higher than the test control group’s 38.62, $p = .02, d = .72$. Other between-group differences did not reach significance.

<table>
<thead>
<tr>
<th>Table 30: Adjusted Means for 3rd Person -s in EI Posttests</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Posttest 1</strong></td>
</tr>
<tr>
<td>Mean</td>
</tr>
<tr>
<td>CR group ($n=32$)</td>
</tr>
<tr>
<td>IR group ($n=32$)</td>
</tr>
<tr>
<td>TkC group ($n=22$)</td>
</tr>
<tr>
<td>TtC group ($n=21$)</td>
</tr>
</tbody>
</table>

Note. Covariates were evaluated at the value of 36.49.
CR = corrective recast; IR = implicit recast; TkC = task control; TtC = test control.

An ANCOVA for EI posttest 2 controlling for the effect of pretest scores found a significant between-group difference, $F(3, 102) = 5.24, p = .00, \eta^2_p = .13$. Follow-up tests showed that both the CR group’s adjusted mean 52.55 and the IR group’s 52.57 were significantly higher than the test control group’s 37.53, with $p = .00, d = .86$ for both of the comparisons. Other between-group differences were not significant in posttest 2 (see Table 31).
### Table 31: A Summary of within- and between-Subject Differences in EI Tests for 3rd Person -s

<table>
<thead>
<tr>
<th></th>
<th>df</th>
<th>p value</th>
<th>Effect size (d)</th>
<th>Effect size ($\eta^2$)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Within-subject</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CR group: pretest - posttest 1</td>
<td>31</td>
<td>.00***</td>
<td>.96*</td>
<td></td>
</tr>
<tr>
<td>CR group: pretest - posttest 2</td>
<td>31</td>
<td>.00***</td>
<td>1.14*</td>
<td></td>
</tr>
<tr>
<td>IR group: pretest - posttest 1</td>
<td>31</td>
<td>.00***</td>
<td>.92*</td>
<td></td>
</tr>
<tr>
<td>IR group: pretest - posttest 2</td>
<td>31</td>
<td>.00***</td>
<td>1.02*</td>
<td></td>
</tr>
<tr>
<td>TkC group: pretest - posttest 1</td>
<td>21</td>
<td>.02*</td>
<td>.66</td>
<td></td>
</tr>
<tr>
<td>TkC group: pretest - posttest 2</td>
<td>21</td>
<td>.00***</td>
<td>.84*</td>
<td></td>
</tr>
<tr>
<td>TkC group: omnibus</td>
<td>2</td>
<td>.91</td>
<td>.01</td>
<td></td>
</tr>
<tr>
<td><strong>Between-subject</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pretest: CR - IR</td>
<td>62</td>
<td>.99</td>
<td>.36</td>
<td></td>
</tr>
<tr>
<td>Pretest: CR - TkC</td>
<td>52</td>
<td>1.00</td>
<td>.16</td>
<td></td>
</tr>
<tr>
<td>Pretest: CR - TtC</td>
<td>51</td>
<td>.23</td>
<td>.58</td>
<td></td>
</tr>
<tr>
<td>Pretest: IR - TkC</td>
<td>52</td>
<td>.39</td>
<td>.52</td>
<td></td>
</tr>
<tr>
<td>Pretest: IR - TtC</td>
<td>51</td>
<td>1.00</td>
<td>.25</td>
<td></td>
</tr>
<tr>
<td>Pretest: TkC - TtC</td>
<td>41</td>
<td>.09</td>
<td>.72</td>
<td></td>
</tr>
<tr>
<td>Posttest 1: CR - IR</td>
<td>62</td>
<td>1.00</td>
<td>.15</td>
<td></td>
</tr>
<tr>
<td>Posttest 1: CR - TkC</td>
<td>52</td>
<td>1.00</td>
<td>.31</td>
<td></td>
</tr>
<tr>
<td>Posttest 1: CR - TtC</td>
<td>51</td>
<td>.02*</td>
<td>.72</td>
<td></td>
</tr>
<tr>
<td>Posttest 1: IR - TkC</td>
<td>52</td>
<td>1.00</td>
<td>.17</td>
<td></td>
</tr>
<tr>
<td>Posttest 1: IR - TtC</td>
<td>51</td>
<td>.11</td>
<td>.57</td>
<td></td>
</tr>
<tr>
<td>Posttest 1: TkC - TtC</td>
<td>41</td>
<td>.80</td>
<td>.40</td>
<td></td>
</tr>
<tr>
<td>Posttest 2: CR - IR</td>
<td>62</td>
<td>1.00</td>
<td>.00</td>
<td></td>
</tr>
<tr>
<td>Posttest 2: CR - TkC</td>
<td>52</td>
<td>1.00</td>
<td>.25</td>
<td></td>
</tr>
<tr>
<td>Posttest 2: CR - TtC</td>
<td>51</td>
<td>.00***</td>
<td>.86*</td>
<td></td>
</tr>
<tr>
<td>Posttest 2: IR - TkC</td>
<td>52</td>
<td>1.00</td>
<td>.26</td>
<td></td>
</tr>
<tr>
<td>Posttest 2: IR - TtC</td>
<td>51</td>
<td>.00***</td>
<td>.86*</td>
<td></td>
</tr>
<tr>
<td>Posttest 2: TkC - TtC</td>
<td>41</td>
<td>.16</td>
<td>.60</td>
<td></td>
</tr>
</tbody>
</table>

Note. CR = corrective recast; IR = implicit recast; TkC = task control; TtC = test control.  
* p ≤ .05.  ** p ≤ .01.  *** p ≤ .001.  # = large effect size.

**A Summary of within- and between-subject differences in EI tests for 3rd person -s**

- The CR, IR and task control groups significantly improved from EI pretest to
posttests, but the test control group did not.

- The CR group was significantly more accurate in production of 3rd person -s than the test control group, but not than other groups in EI posttest 1.
- Both the CR and IR groups significantly outperformed the test control group, but not the task control group in EI posttest 2.
- The CF groups did not differ significantly in EI posttests for 3rd person -s.

5.2.2 Results for 3rd person -s: Oral production

Learners’ accuracy in the production of 3rd person -s was also measured by means of oral production tests. The OP test results are presented in the following section.

Descriptive statistics

As Table 32 and Figure 4 illustrate, the CR group had a mean score of 40.62 in accuracy of 3rd person -s in the OP pretest, which rose to 66.19 in posttest 1 and then shrank to 52.22 in posttest 2. The IR group’s mean also increased from pretest (M = 55.04) to posttest 1 (M = 61.79) and then edged down in posttest 2 (M = 58.77). The task control group gained a mean score of 47.12 in the pretest, and the mean went up to 58.85 in posttest 1 but slipped back to 45.68 in posttest 2, which was even lower than its pretest mean. The test control group decreased from pretest (M = 54.56) to posttest 1 (M= 49.28) and then to posttest 2 (M = 44.81).

<table>
<thead>
<tr>
<th></th>
<th>Pretest</th>
<th></th>
<th>Posttest 1</th>
<th></th>
<th>Posttest 2</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>CR group (n=27)</td>
<td>40.62</td>
<td>19.87</td>
<td>66.19</td>
<td>21.17</td>
<td>52.22</td>
<td>23.00</td>
</tr>
<tr>
<td>IR group (n=30)</td>
<td>55.04</td>
<td>17.61</td>
<td>61.79</td>
<td>23.37</td>
<td>58.77</td>
<td>20.83</td>
</tr>
<tr>
<td>TkC group (n=19)</td>
<td>47.12</td>
<td>23.61</td>
<td>58.85</td>
<td>26.63</td>
<td>45.68</td>
<td>21.01</td>
</tr>
<tr>
<td>TtC group (n=19)</td>
<td>54.56</td>
<td>23.61</td>
<td>49.28</td>
<td>18.55</td>
<td>44.81</td>
<td>21.92</td>
</tr>
<tr>
<td>Total (n=95)</td>
<td>49.26</td>
<td>21.38</td>
<td>59.95</td>
<td>23.01</td>
<td>51.50</td>
<td>22.13</td>
</tr>
</tbody>
</table>

Note. CR = corrective recast; IR = implicit recast; TkC = task control; TtC = test control.
Figure 4: Means for 3rd Person -s in OP Tests over Time by Group

Inferential statistics

Omnibus results

A mixed 3 x 4 ANOVA across groups over time of OP scores for 3rd person -s revealed that the interaction between time and group was significant, $F(6, 182) = 6.29, p = .00, \eta^2_p = .17$. The significant interaction was further analyzed by examining the main effect of time in each group and the main effect of group at each testing point.

Results for the effect of time in each group

The CR group’s scores for accuracy of 3rd person -s in OP tests were compared among testing times using a non-parametric repeated-measures ANOVA as its scores in posttest 1 did not have a normal distribution. The Friedman test indicated a significant main effect for time, $\chi^2(2) = 26.74, p = .00$. Wilcoxon signed-ranks or paired-samples $t$ tests as a post-hoc procedure were conducted with a Bonferroni correction applied, resulting in the alpha level
set at $p < .02$. The CR group had a significant improvement from pretest to posttest 1, $Z = 4.04$, $p = .00$, $d = 1.14$, and then a significant decrease from posttest 1 to posttest 2, $Z = 3.27$, $p = .00$, $d = .69$, but its mean score in posttest 2 was still significantly higher than in the pretest, $t(26) = 3.16$, $p = .00$, $d = .61$.

The IR group’s OP scores were analyzed by means of a repeated-measures ANOVA. The analysis found no significant difference over time, $F(2, 58) = 2.24$, $p > .05$, $\eta^2_p = .07$.

A repeated-measures ANOVA for the task control group indicated a significant main effect for time, $F(2, 36) = 7.23$, $p = .00$, $\eta^2_p = .29$. Post-hoc analyses showed the task control group’s accuracy of 3rd person -s improved significantly from pretest to posttest 1, $p = .04$, $d = .65$, and then declined significantly from posttest 1 to posttest 2, $p = .01$, $d = .91$, so that there was no significant difference between pretest and posttest 2.

A repeated-measures ANOVA for the test control group showed no significant effect for time, $F(2, 36) = 2.79$, $p > .05$, $\eta^2_p = .13$.

**Results for the effect of group at each testing point**

Participants’ scores for accuracy of 3rd person -s in the OP pretest were compared across groups by means of a one-way ANOVA. The analysis showed a significant effect for group, $F(3, 91) = 2.81$, $p = .04$, $\eta^2 = .08$. Post-hoc comparisons indicated that the CR - IR difference approached significance, $p = .06$, $d = .77$. Consequently, OP posttest scores were compared using an ANCOVA or a Quade (1967) non-parametric ANCOVA (when the data did not have a normal distribution), which was carried out in SPSS by ranking the variables, running a linear regression of the ranks of the dependent variable on the ranks of the covariate and then running a one-way ANOVA of the residuals produced in the prior step across groups.

A Quade ANCOVA of the scores in posttest 1 determined that there was a significant between-group difference, $F(3, 91) = 6.83$, $p = .00$, $\eta^2 = .18$. Bonferroni post-hoc procedures showed that the CR group with an unstandardized residual mean of 12.91 performed significantly better than the test control group with a residual mean of -17.26, $p = .00$, $d = 1.46$. Although the IR and the task control groups’ scores were not significantly higher than the test control group’s, there was a large effect size, $d = .83$ for the IR - test control difference, and $d = .86$ for the task control - test control difference.
Table 33: Adjusted Means for 3rd Person -s in OP Posttests

<table>
<thead>
<tr>
<th>Group</th>
<th>Posttest 1</th>
<th>Posttest 2</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean a</td>
<td>SD</td>
<td>Mean b</td>
<td>Std Error</td>
</tr>
<tr>
<td>CR group (n=27)</td>
<td>12.91</td>
<td>24.19</td>
<td>58.02</td>
<td>3.30</td>
</tr>
<tr>
<td>IR group (n=30)</td>
<td>-.83</td>
<td>22.86</td>
<td>54.89</td>
<td>3.09</td>
</tr>
<tr>
<td>TkC group (n=19)</td>
<td>.22</td>
<td>23.60</td>
<td>47.12</td>
<td>3.84</td>
</tr>
<tr>
<td>TtC group (n=19)</td>
<td>-17.26</td>
<td>16.35</td>
<td>41.25</td>
<td>3.86</td>
</tr>
</tbody>
</table>

Note. a. Means of unstandardized residuals were compared in a non-parametric ANCOVA.
   b. Covariates were evaluated at the value of 49.26.
   CR = corrective recast; IR = implicit recast; TkC = task control; TtC = test control.

An ANCOVA was carried out for OP posttest 2 controlling for the effect of pretest scores. The analysis determined that there was a significant between-group difference among adjusted means, $F(3, 90) = 4.50$, $p = .01$, $\eta_p^2 = .13$. Bonferroni post-hoc analyses revealed that the CR group with an adjusted mean of 58.02 ($p = .01, d = .78$) and the IR group with an adjusted mean of 54.89 ($p = .04, d = .64$) performed significantly better than the test control group with an adjusted mean of 41.25, but not significantly better than the task control group’s 47.12. No significant difference was observed between the CF groups (see Table 34).
Table 34: A Summary of within- and between-Subject Differences in OP Tests for 3rd Person -s

<table>
<thead>
<tr>
<th></th>
<th>df</th>
<th>p value</th>
<th>Effect size (d)</th>
<th>Effect size (η²)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Within-subject</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CR group: pretest -</td>
<td>26</td>
<td>.00***</td>
<td>1.14#</td>
<td></td>
</tr>
<tr>
<td>posttest 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CR group: pretest -</td>
<td>26</td>
<td>.00***</td>
<td>.61</td>
<td></td>
</tr>
<tr>
<td>posttest 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CR group: posttest1</td>
<td>26</td>
<td>.00***</td>
<td>.69</td>
<td></td>
</tr>
<tr>
<td>- posttest 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IR group: omnibus</td>
<td>2</td>
<td>.12</td>
<td></td>
<td>.07</td>
</tr>
<tr>
<td>TkC group: pretest -</td>
<td>18</td>
<td>.04*</td>
<td>.65</td>
<td></td>
</tr>
<tr>
<td>posttest 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TkC group: pretest -</td>
<td>18</td>
<td>1.00</td>
<td>.09</td>
<td></td>
</tr>
<tr>
<td>posttest 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TkC group: posttest1</td>
<td>18</td>
<td>.01**</td>
<td>.91#</td>
<td></td>
</tr>
<tr>
<td>- posttest 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TkC group: omnibus</td>
<td>2</td>
<td>.08</td>
<td>.13</td>
<td></td>
</tr>
<tr>
<td><strong>Between-subject</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pretest: CR - IR</td>
<td>55</td>
<td>.06</td>
<td>.77</td>
<td></td>
</tr>
<tr>
<td>Pretest: CR - TkC</td>
<td>44</td>
<td>1.00</td>
<td>.30</td>
<td></td>
</tr>
<tr>
<td>Pretest: CR - TtC</td>
<td>44</td>
<td>.17</td>
<td>.64</td>
<td></td>
</tr>
<tr>
<td>Pretest: IR - TkC</td>
<td>47</td>
<td>1.00</td>
<td>.38</td>
<td></td>
</tr>
<tr>
<td>Pretest: IR - TtC</td>
<td>47</td>
<td>1.00</td>
<td>.02</td>
<td></td>
</tr>
<tr>
<td>Pretest: TkC - TtC</td>
<td>36</td>
<td>1.00</td>
<td>.32</td>
<td></td>
</tr>
<tr>
<td>Posttest 1: CR - IR</td>
<td>55</td>
<td>.13</td>
<td>.58</td>
<td></td>
</tr>
<tr>
<td>Posttest 1: CR - TkC</td>
<td>44</td>
<td>.36</td>
<td>.53</td>
<td></td>
</tr>
<tr>
<td>Posttest 1: CR - TtC</td>
<td>44</td>
<td>.00***</td>
<td>1.46#</td>
<td></td>
</tr>
<tr>
<td>Posttest 1: IR - TkC</td>
<td>47</td>
<td>1.00</td>
<td>.05</td>
<td></td>
</tr>
<tr>
<td>Posttest 1: IR - TtC</td>
<td>47</td>
<td>.08</td>
<td>.83#</td>
<td></td>
</tr>
<tr>
<td>Posttest 1: TkC - TtC</td>
<td>36</td>
<td>.11</td>
<td>.86#</td>
<td></td>
</tr>
<tr>
<td>Posttest 2: CR - IR</td>
<td>55</td>
<td>1.00</td>
<td>.15</td>
<td></td>
</tr>
<tr>
<td>Posttest 2: CR - TkC</td>
<td>44</td>
<td>.20</td>
<td>.51</td>
<td></td>
</tr>
<tr>
<td>Posttest 2: CR - TtC</td>
<td>44</td>
<td>.01**</td>
<td>.78</td>
<td></td>
</tr>
<tr>
<td>Posttest 2: IR - TkC</td>
<td>47</td>
<td>.72</td>
<td>.36</td>
<td></td>
</tr>
<tr>
<td>Posttest 2: IR - TtC</td>
<td>47</td>
<td>.04*</td>
<td>.64</td>
<td></td>
</tr>
<tr>
<td>Posttest 2: TkC - TtC</td>
<td>36</td>
<td>1.00</td>
<td>.27</td>
<td></td>
</tr>
</tbody>
</table>

Note. CR = corrective recast; IR = implicit recast; TkC = task control; TtC = test control.
* p ≤ .05.  ** p ≤ .01.  *** p ≤ .001.  # = large effect size.

A summary of within- and between-subject differences in OP tests for 3rd person -s

- The CR group significantly improved their accuracy for 3rd person -s from OP pretest to posttest 1 and sustained the improvement in posttest 2 although a significant
decline occurred between posttests.

- The task control group also performed significantly better in posttest 1 than in the pretest but the improvement disappeared in posttest 2.

- The IR and test control groups did not change significantly from pretest to posttests.

- There was a significant effect for group in OP scores for 3rd person -s prior to the treatment.

- The CR group performed significantly better than the test control group in OP posttests.

- The IR group scored much higher than the test control group with a large effect size in posttest 1 and significantly in posttest 2.

- The task control group also outperformed the test control group with a large effect size in posttest 1, but not in posttest 2.

- There was no significant difference between the CF groups in OP posttests, but the CR-IR difference had a medium effect size in OP posttest 1.

5.2.3 Results for embedded questions: Elicited imitation

Descriptive statistics

Means and SDs of the participants’ scores for embedded questions in the elicited imitation test are summarized in Table 35. The CR group’s mean scores rose quickly from pretest (M = 50.30) to posttest 1 (M = 64.25) and got even higher in posttest 2 (M = 68.14) as illustrated by Figure 5. The IR group’s scores also increased from pretest (M = 56.40) to posttest 1 (M = 63.17) and the improvement sustained in posttest 2 (M = 64.13). The task control group’s scores did not change much from pretest (M = 56.36) to posttest 1 (M = 57.30) but had a dramatic growth from posttest 1 to posttest 2 (M = 70.42). In contrast, the test control group’s scores edged down from pretest (M = 64.38) to posttest 1 (M = 61.61) and then recovered slightly in posttest 2 (M = 62.84).
Table 35: Descriptive Statistics of Scores for Embedded Questions in EI Tests

<table>
<thead>
<tr>
<th></th>
<th>Pretest</th>
<th>Posttest 1</th>
<th>Posttest 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
</tr>
<tr>
<td>CR group (n=26)</td>
<td>50.30</td>
<td>12.09</td>
<td>64.25</td>
</tr>
<tr>
<td>IR group (n=31)</td>
<td>56.40</td>
<td>15.52</td>
<td>63.17</td>
</tr>
<tr>
<td>TkC group (n=20)</td>
<td>56.36</td>
<td>16.82</td>
<td>57.30</td>
</tr>
<tr>
<td>TtC group (n=18)</td>
<td>64.38</td>
<td>13.62</td>
<td>61.61</td>
</tr>
<tr>
<td>Total (n=95)</td>
<td>56.23</td>
<td>15.13</td>
<td>61.93</td>
</tr>
</tbody>
</table>

Note. CR = corrective recast; IR = implicit recast; TkC = task control; TtC = test control.

Figure 5: Means for Embedded Questions in EI Tests over Time by Group

Inferential statistics

Omnibus results

A mixed ANOVA of the EI scores for embedded questions showed that the interaction between time and group was significant, $F(6, 182) = 4.94$, $p = .00$, $\eta^2_p = .14$. The significant interaction was examined by means of the main effect of time in each group and the main effect of group at each testing point.
Results for the effect of time in each group

A repeated-measures ANOVA of the CR group’s scores in EI tests for embedded questions revealed a significant difference over time: \( F (2, 50) = 17.90, p = .00, \eta^2_p = .42 \). Post-hoc analyses showed the CR group’s scores in posttest 1 \((p = .00, d = 1.06)\) and posttest 2 \((p = .00, d = 1.04)\) were both significantly higher than in the pretest, but there was no significant difference between posttests.

A repeated-measures ANOVA for the IR group found a significant effect for time, \( F (2, 60) = 5.26, p = .01, \eta^2_p = .15 \). Post-hoc comparisons illustrated that the difference between pretest and posttest 1 did not reach significance, but the difference between pretest and posttest 2 was significant, \( p = .02, d = .55 \). No significant difference was observed between posttests.

A repeated-measures ANOVA for the task control group also indicated a significant difference over time, \( F (2, 38) = 9.62, p = .00, \eta^2_p = .34 \). Post-hoc analyses showed the task control group did not improve significantly from pretest to posttest 1, but had a significant growth between posttests, \( p = .01, d = .75 \), so that the scores in posttest 2 were also significantly higher than in the pretest, \( p = .00, d = .91 \).

A repeated-measures ANOVA for the test control group revealed no significant difference between testing times, \( F (2, 34) = .38, p > .05, \eta^2_p = .02 \).

Results for the effect of group at each testing point

A one-way ANOVA of EI pretest scores determined that there was a significant between-group difference in accuracy of embedded questions prior to the treatment, \( F (3, 91) = 3.30, p = .02, \eta^2 = .10 \). Bonferroni post-hoc comparisons showed that the CR group had a significant lower mean than the test control group, \( p = .01, d = 1.09 \). Other between-group differences were not significant.

Because of the significant pre-treatment difference between groups, EI posttest scores were compared using ANCOVAs controlling for the effect of pretest scores. The ANCOVA for posttest 1 showed a significant between-group difference among adjusted means, \( F (3, 90) = 3.34, p = .02, \eta^2_p = .10 \). Post-hoc procedures determined that the CR group’s adjusted mean 67.18 was significantly higher than the task control group’s 57.23, \( p = .04, d = .73 \). Other between-group differences were not significant in EI posttest 1.
Table 36: Adjusted Means for Embedded Questions in EI Posttests

<table>
<thead>
<tr>
<th></th>
<th>Posttest 1</th>
<th></th>
<th>Posttest 2</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Std Error</td>
<td>Mean</td>
<td>Std Error</td>
</tr>
<tr>
<td>CR group (n=26)</td>
<td>67.18</td>
<td>2.40</td>
<td>71.55</td>
<td>2.92</td>
</tr>
<tr>
<td>IR group (n=31)</td>
<td>63.09</td>
<td>2.15</td>
<td>64.03</td>
<td>2.61</td>
</tr>
<tr>
<td>TkC group (n= 20)</td>
<td>57.23</td>
<td>2.68</td>
<td>70.34</td>
<td>3.25</td>
</tr>
<tr>
<td>TtC group (n= 18)</td>
<td>57.59</td>
<td>2.91</td>
<td>58.15</td>
<td>3.53</td>
</tr>
</tbody>
</table>

Note. Covariates were evaluated at the value of 56.23.
CR = corrective recast; IR = implicit recast; TkC = task control; TtC = test control.

The ANCOVA for EI posttest 2 found a significant between-group difference, $F (3, 90) = 3.49, p = .02, \eta^2_p = .10$. Follow-up comparisons showed that the CR group with an adjusted mean score of 71.55 significantly outperformed the test control group (adjusted M = 58.15), $p = .03, d = .82$. Other between-group differences were not significant in posttest 2 (see Table 37).
### Table 37: A Summary of within- and between-Subject Differences in EI Tests for Embedded Questions

<table>
<thead>
<tr>
<th></th>
<th>df</th>
<th>p value</th>
<th>Effect size (d)</th>
<th>Effect size ($\eta^2$)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Within-subject</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CR group: pretest - posttest 1</td>
<td>25</td>
<td>.00***</td>
<td>1.06*</td>
<td></td>
</tr>
<tr>
<td>CR group: pretest - posttest 2</td>
<td>25</td>
<td>.00***</td>
<td>1.04*</td>
<td></td>
</tr>
<tr>
<td>CR group: posttest 1 - posttest 2</td>
<td>25</td>
<td>.74</td>
<td>.24</td>
<td></td>
</tr>
<tr>
<td>IR group: pretest - posttest 1</td>
<td>30</td>
<td>.08</td>
<td>.42</td>
<td></td>
</tr>
<tr>
<td>IR group: pretest - posttest 2</td>
<td>30</td>
<td>.02*</td>
<td>.55</td>
<td></td>
</tr>
<tr>
<td>IR group: posttest 1 - posttest 2</td>
<td>30</td>
<td>1.00</td>
<td>.08</td>
<td></td>
</tr>
<tr>
<td>TkC group: pretest - posttest 1</td>
<td>19</td>
<td>1.00</td>
<td>.07</td>
<td></td>
</tr>
<tr>
<td>TkC group: pretest - posttest 2</td>
<td>19</td>
<td>.00***</td>
<td>.91*</td>
<td></td>
</tr>
<tr>
<td>TkC group: posttest 1 - posttest 2</td>
<td>19</td>
<td>.01**</td>
<td>.75</td>
<td></td>
</tr>
<tr>
<td>TtC group: omnibus</td>
<td>2</td>
<td>.69</td>
<td>.02</td>
<td></td>
</tr>
<tr>
<td>Pretest: CR - IR</td>
<td>55</td>
<td>.72</td>
<td>.44</td>
<td></td>
</tr>
<tr>
<td>Pretest: CR - TkC</td>
<td>44</td>
<td>1.00</td>
<td>.42</td>
<td></td>
</tr>
<tr>
<td>Pretest: CR - TtC</td>
<td>42</td>
<td>.01**</td>
<td>1.09*</td>
<td></td>
</tr>
<tr>
<td>Pretest: IR - TkC</td>
<td>49</td>
<td>1.00</td>
<td>.00</td>
<td></td>
</tr>
<tr>
<td>Pretest: IR - TtC</td>
<td>47</td>
<td>.41</td>
<td>.55</td>
<td></td>
</tr>
<tr>
<td>Pretest: TkC - TtC</td>
<td>36</td>
<td>.57</td>
<td>.52</td>
<td></td>
</tr>
<tr>
<td>Posttest 1: CR - IR</td>
<td>55</td>
<td>1.00</td>
<td>.30</td>
<td></td>
</tr>
<tr>
<td>Posttest 1: CR - TkC</td>
<td>44</td>
<td>.04*</td>
<td>.73</td>
<td></td>
</tr>
<tr>
<td>Posttest 1: CR - TtC</td>
<td>42</td>
<td>.09</td>
<td>.70</td>
<td></td>
</tr>
<tr>
<td>Posttest 1: IR - TkC</td>
<td>49</td>
<td>.55</td>
<td>.43</td>
<td></td>
</tr>
<tr>
<td>Posttest 1: IR - TtC</td>
<td>47</td>
<td>.79</td>
<td>.40</td>
<td></td>
</tr>
<tr>
<td>Posttest 1: TkC - TtC</td>
<td>36</td>
<td>1.00</td>
<td>.03</td>
<td></td>
</tr>
<tr>
<td>Posttest 2: CR - IR</td>
<td>55</td>
<td>.35</td>
<td>.46</td>
<td></td>
</tr>
<tr>
<td>Posttest 2: CR - TkC</td>
<td>44</td>
<td>1.00</td>
<td>.07</td>
<td></td>
</tr>
<tr>
<td>Posttest 2: CR - TtC</td>
<td>42</td>
<td>.03*</td>
<td>.82*</td>
<td></td>
</tr>
<tr>
<td>Posttest 2: IR - TkC</td>
<td>49</td>
<td>.80</td>
<td>.39</td>
<td></td>
</tr>
<tr>
<td>Posttest 2: IR - TtC</td>
<td>47</td>
<td>1.00</td>
<td>.36</td>
<td></td>
</tr>
<tr>
<td>Posttest 2: TkC - TtC</td>
<td>36</td>
<td>.08</td>
<td>.74</td>
<td></td>
</tr>
</tbody>
</table>

**Note.** CR = corrective recast; IR = implicit recast; TkC group = task control; TtC = test control.  
* p≤ .05.  ** p≤ .01.  *** p≤ .001.  # = large effect size.

---

**A summary of within- and between-subject differences in EI tests for embedded questions**

- The CR group significantly improved their accuracy in production of embedded
questions in EI posttests.

• The IR and task control groups did not have a significant improvement in accuracy of embedded questions until posttest 2.

• The test control group did not change significantly from EI pretest to posttests.

• There was a significant between-group difference in EI pretest for embedded questions.

• The CR group significantly outperformed the task control group in EI posttest 1, but the IR group did not.

• The CR group was significantly more accurate in production of embedded questions than the test control group in EI posttest 2, but the IR group did not.

• There was no significant difference between the two CF groups in EI posttests for embedded questions.

5.2.4 Results for embedded questions: Oral production

Descriptive statistics

Table 38 presents each group’s means and SDs of scores in the oral production tests for embedded questions. The CR group had a mean score of 69.48 in the pretest and gained a higher mean of 87.27 in posttest 1 and an even higher mean of 92.95 in posttest 2 (see Figure 6). The IR group’s scores also increased from 67.91 in the pretest to 84.95 in posttest 1 and then decreased slightly to 81.14 in posttest 2. The task control group’s means rose all the way from pretest (M = 71.52) to posttest 1 (M = 82.99) and then to posttest 2 (M = 86.77). The test control group’s means fluctuated from pretest (M = 75.66) down to pretest 1 (M = 70.44) and then up to posttest 2 (M = 82.71).
Table 38: Descriptive Statistics of Scores for Embedded Questions in OP Tests

<table>
<thead>
<tr>
<th></th>
<th>Pretest</th>
<th></th>
<th></th>
<th>Posttest 1</th>
<th></th>
<th></th>
<th>Posttest 2</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>CR group (n=13)</td>
<td>69.48</td>
<td>9.12</td>
<td>87.27</td>
<td>13.84</td>
<td>92.95</td>
<td>7.63</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IR group (n=18)</td>
<td>67.91</td>
<td>18.24</td>
<td>84.95</td>
<td>12.57</td>
<td>81.14</td>
<td>15.13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TkC group (n=13)</td>
<td>71.52</td>
<td>12.41</td>
<td>82.99</td>
<td>18.22</td>
<td>86.77</td>
<td>11.84</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TtC group (n=11)</td>
<td>75.66</td>
<td>12.65</td>
<td>70.44</td>
<td>11.49</td>
<td>82.71</td>
<td>10.80</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total (n=55)</td>
<td>70.68</td>
<td>13.97</td>
<td>82.13</td>
<td>15.09</td>
<td>85.58</td>
<td>12.63</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. CR = corrective recast; IR = implicit recast; TkC = task control; TtC = test control.

Figure 6: Means for Embedded Questions in OP Tests over Time by Group

Inferential statistics

Omnibus results

A mixed ANOVA of the scores for accuracy of embedded questions in OP tests showed that the time x group interaction was significant, $F (6, 102) = 3.30, p = .01, \eta^2 = .16$. 
Results for the effect of time in each group

A repeated-measures ANOVA was carried out for the CR group with a Greenhouse-Geisser correction ($\varepsilon = .70$) because the assumption of sphericity had been violated as determined by the Mauchly’s test, $\chi^2 (2) = 6.17, p = .05$. The analysis showed a significant difference over time, $F (1.40, 16.79) = 21.35, p = .00, \eta^2_p = .64$. Bonferroni post-hoc comparisons revealed the CR group significantly improved its accuracy in the production of embedded questions from pretest to posttest 1, $p = .00, d = 1.35$, and from pretest to posttest 2, $p = .00, d = 2.43$. The CR group’s difference between OP posttests was not significant.

A repeated-measures ANOVA for the IR group found a significant main effect for time, $F (2, 34) = 8.64, p = .00, \eta^2_p = .34$. Bonferroni post-hoc analyses indicated the IR group’s scores in posttest 1 ($p = .00, d = 1.04$) and posttest 2 ($p = .01, d = .73$) were both significantly higher than in the pretest. There was no significant difference between OP posttests.

A repeated-measures ANOVA for the task control group also indicated a significant main effect for time, $F (2, 24) = 6.81, p = .01, \eta^2_p = .36$. Post-hoc analyses showed the task control group performed significantly better in posttest 1 ($p = .03, d = .87$) and posttest 2 ($p = .02, d = .87$) than in the pretest. Their scores did not change significantly from posttest 1 to posttest 2.

A non-parametric repeated-measures ANOVA was conducted for the test control group as its OP scores in posttest 2 did not have a normal distribution. The Friedman test showed the test control group’s improvement in accuracy of embedded questions was not significant, $\chi^2 (2) = 4.91, p > .05$.

Results for the effect of group at each testing point

A Welch ANOVA was conducted for OP pretest for embedded questions as the assumption of homogeneity of variances was violated. The analysis indicated that the four groups did not differ significantly prior to the treatment when measured by means of OP tests, $F (3, 27.09) = .76, p > .05, \eta^2 = .04$.

A one-way ANOVA of scores in OP posttest 1 showed a significant effect for group, $F (3, 51) = 3.31, p = .03, \eta^2 = .16$. Bonferroni post-hoc comparisons indicated that the CR group significantly outperformed the test control group, $p = .03, d = 1.32$ but not other groups. The
IR and task control group did not significantly outperform the test control group in OP posttest 1, but the IR - test control difference had a large effect size, $d = 1.21$, and the difference between the control groups also had a large effect size, $d = .82$.

A Kruskal-Wallis test was carried out for OP posttest 2 as the test control group’s scores did not have a normal distribution. The non-parametric test revealed between-group differences approached significance, $\chi^2 (3) = 6.81, p = .08$. Pairwise comparisons were conducted with a Bonferroni correction, resulting in the alpha level set at $p < .01$. The CR group performed significantly better than the IR group, $t (26.44) = 2.85, p = .01, d = .99$, and also better than the test control group with a large effect size ($Z = 2.33, p = .02, d = 1.10$). The IR group did not significantly differ from any control group, and the difference between the control groups was not significant either (see Table 39).
Table 39: A Summary of within- and between-Subject Differences in OP Tests for Embedded Questions

<table>
<thead>
<tr>
<th></th>
<th>$df$</th>
<th>$p$ value</th>
<th>Effect size $(d)$</th>
<th>Effect size $(\eta^2)$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Within-subject</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CR group: pretest - pretest 1</td>
<td>12</td>
<td>.00*</td>
<td>1.35#</td>
<td></td>
</tr>
<tr>
<td>CR group: pretest - pretest 2</td>
<td>12</td>
<td>.00***</td>
<td>2.43#</td>
<td></td>
</tr>
<tr>
<td>CR group: pretest - posttest 1</td>
<td>12</td>
<td>.72</td>
<td>.39</td>
<td></td>
</tr>
<tr>
<td>IR group: pretest - pretest 1</td>
<td>17</td>
<td>.00*</td>
<td>1.04#</td>
<td></td>
</tr>
<tr>
<td>IR group: pretest - pretest 2</td>
<td>17</td>
<td>.01*</td>
<td>.73</td>
<td></td>
</tr>
<tr>
<td>IR group: pretest - posttest 2</td>
<td>17</td>
<td>1.00</td>
<td>.18</td>
<td></td>
</tr>
<tr>
<td>TkC group: pretest - pretest 1</td>
<td>12</td>
<td>.03*</td>
<td>.87#</td>
<td></td>
</tr>
<tr>
<td>TkC group: pretest - pretest 2</td>
<td>12</td>
<td>.02*</td>
<td>.87#</td>
<td></td>
</tr>
<tr>
<td>TkC group: posttest 1 - posttest 2</td>
<td>12</td>
<td>1.00</td>
<td>.26</td>
<td></td>
</tr>
<tr>
<td>TkC group: omnibus</td>
<td></td>
<td>.09</td>
<td></td>
<td>.06</td>
</tr>
<tr>
<td>Pretest: omnibus</td>
<td>3</td>
<td>.53</td>
<td></td>
<td>.04</td>
</tr>
<tr>
<td>Posttest 1: CR - IR</td>
<td>29</td>
<td>1.00</td>
<td>.18</td>
<td></td>
</tr>
<tr>
<td>Posttest 1: CR - TkC</td>
<td>24</td>
<td>1.00</td>
<td>.26</td>
<td></td>
</tr>
<tr>
<td>Posttest 1: CR - TtC</td>
<td>22</td>
<td>.03*</td>
<td>1.32#</td>
<td></td>
</tr>
<tr>
<td>Posttest 1: IR - TkC</td>
<td>29</td>
<td>1.00</td>
<td>.13</td>
<td></td>
</tr>
<tr>
<td>Posttest 1: IR - TtC</td>
<td>27</td>
<td>.06</td>
<td>1.21#</td>
<td></td>
</tr>
<tr>
<td>Posttest 2: TkC - TtC</td>
<td>22</td>
<td>.22</td>
<td>.82#</td>
<td></td>
</tr>
<tr>
<td>Posttest 2: CR - IR</td>
<td>29</td>
<td>.01**</td>
<td>.99#</td>
<td></td>
</tr>
<tr>
<td>Posttest 2: CR - TkC</td>
<td>24</td>
<td>.13</td>
<td>.62</td>
<td></td>
</tr>
<tr>
<td>Posttest 2: CR - TtC</td>
<td>22</td>
<td>.02*</td>
<td>1.10#</td>
<td></td>
</tr>
<tr>
<td>Posttest 2: IR - TkC</td>
<td>29</td>
<td>.27</td>
<td>.41</td>
<td></td>
</tr>
<tr>
<td>Posttest 2: IR - TtC</td>
<td>27</td>
<td>.98</td>
<td>.12</td>
<td></td>
</tr>
<tr>
<td>Posttest 2: TkC - TtC</td>
<td>22</td>
<td>.54</td>
<td>.36</td>
<td></td>
</tr>
</tbody>
</table>

Note. CR = corrective recast; IR = implicit recast; TkC = task control; TtC = test control. * $p \leq .05$. ** $p \leq .01$. *** $p \leq .001$. # = large effect size.

A summary of within- and between-subject differences in OP tests for embedded questions

- The CR, IR and task control groups all significantly improved their accuracy of embedded questions from OP pretest to posttest 1 and maintained their improvement in posttest 2.

- The test control group did not change significantly from OP pretest to posttests for embedded questions.
The four groups did not differ in accuracy of embedded questions prior to the treatment as measured by means of OP tests.

The CR, IR and task control groups outperformed the test control group with a large effect size in OP posttest 1 for embedded questions.

The CR group scored higher than the IR and test control groups with a large effect size in OP posttest 2 for embedded questions.

5.2.5 Results for reported noticing

The exit questionnaire was conducted to collect information on noticing of explicit/implicit recasts. Both the CR and IR groups had a high level of noticing of the CF on grammar errors, but differed in accuracy of perceiving the target structures. In the CR group \((n = 33)\), 30 participants claimed noticing of CF on grammar errors, 22 recalled that the teacher was correcting a 3rd person \(-s\) error during the quasi-experiment, and 14 recalled correction of an embedded question error. In the IR group \((n = 33)\), 25 participants reported noticing of CF on grammar errors, but only seven recalled correction of a 3rd person \(-s\) error, and 3 recalled correction of an embedded question error.

<table>
<thead>
<tr>
<th></th>
<th>Noticing of CF on grammar</th>
<th>Noticing of CF on 3rd person (-s)</th>
<th>Noticing of CF on embedded questions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Frequency</td>
<td>Percentage</td>
<td>Frequency</td>
</tr>
<tr>
<td>CR group ((n = 33))</td>
<td>30</td>
<td>91%</td>
<td>22</td>
</tr>
<tr>
<td>IR group ((n = 33))</td>
<td>25</td>
<td>76%</td>
<td>7</td>
</tr>
<tr>
<td>Total ((n = 66))</td>
<td>55</td>
<td>83%</td>
<td>29</td>
</tr>
</tbody>
</table>

Note. CR = corrective recast; IR = implicit recast.

Chi-square tests showed that reported noticing of CF on grammar errors was not related to types of corrective feedback (CF), \(\chi^2 (1, 66) = 2.73, p = .10, \phi = .20\), but reported noticing of 3rd person \(-s\) \((\chi^2 (1, 66) = 13.8, p = .00, \phi = .46)\) and embedded questions \((\chi^2 (1, 66) = 9.59,\)
was dependent on the CF condition. The participants were more likely to locate the target structures accurately receiving explicit recasts than receiving implicit recasts, as illustrated by the odds ratios of 7.43 (for 3rd person -s) and 7.34 (embedded questions).

5.3 Discussion

RQ 1 investigates the relative effects of corrective recasts and implicit recasts on the development of implicit knowledge.

Question 1a asks whether corrective recasts affect learners’ implicit knowledge of L2 grammatical features. The answer is affirmative. As shown in the results section, the CR group significantly improved their accuracy in the production of 3rd person -s and embedded questions from pretest to posttest 1 and maintained the significant gains in posttest 2. As for between-group differences, the CR group outperformed not only the test control group in the posttests but also the task control group in EI posttest 1 for embedded questions. This finding corroborates the claim that multiple-move feedback has “double focus on the error” (Erlam & Loewen, 2010, p. 880) and thus is more explicit (see Doughty & Varela, 1998; Loewen & Philp, 2006; Sheen, 2006).

The benefit of corrective recasts could be attributed to their double feedback moves: a repetition and a recast. Firstly, the repetition eased the difficulty of input processing. According to VanPatten’s (2004) Input Processing theory, input is processed for meaning before form because processing capacity is limited. A corrective recast repeated the learner’s erroneous utterance, and thus the repetition may have freed the learner’s attentional resources from meaning and enabled attention to form to occur. Attention to form is necessary for L2 acquisition (Schmidt, 1990; 2001), through which input is filtered into intake for further analysis. Secondly, the rising intonation in repetition rendered the corrective force more explicit. A repetition in a rising tone sounded like an interrogative question and it might have signaled to the learner that negative evidence was being offered. Thirdly, the ensuing recast led to “noticing the gap”. The recast, as positive evidence, was sharply juxtaposed with the erroneous utterance in repetition. The juxtaposition made the target structure more salient to the learner so that he or she could locate the error. The more salient the target structure is made, the more likely that it is noticed and processed further. By comparing in working memory what he uttered and what the teacher provided, the learner might have noticed the gap between his own interlanguage system and the target language. The exit questionnaire provided evidence for this. 91% of the participants in the CR group claimed that they noticed
the teacher’s correction of grammatical errors, 67% of participants recalled that the teacher was correcting a 3rd person -s error, and 42% of participants reported that the teacher was correcting an embedded question error. This “cognitive comparison” (R. Ellis, 1995, p. 90), in turn, pushed grammatical restructuring in long term memory. The finding that explicit recasts were effective in this study, in line with Mifka-Profozic’s (2013) study, which manipulated short partial recasts, lends support to Goo and Mackey’s (2013) claim that recasts work in classroom as well as laboratory settings.

Question 1b asks whether implicit recasts affect learners’ implicit knowledge of L2 grammatical features. This study shows that implicit recasts have limited value in the development of implicit knowledge. The IR group improved their accuracy of 3rd person -s in the EI posttests and accuracy of embedded questions in the EI and OP posttests. However, in the posttests showing the benefit of implicit recasts, the task control group also demonstrated a significant improvement. It can be argued that the significant effect of “interaction + IR” mainly results from oral interaction rather than the CF. This speculation is supported by the evidence that there was no significant difference between the IR and task control groups in all the posttests of implicit knowledge. These findings do not support Long et al.’s (1998) position that implicit recasts facilitate acquisition.

The limited efficacy of implicit recasts can be explained by the ambiguity of the CF, the implicitness of the CF and the redundancy of the target structure. Lyster (1998b) found that recasts as negative evidence and non-corrective repetitions were evenly distributed in content-based lessons, and it was difficult for learners to recognize teachers’ corrective purpose. Although the learners perceived the teacher’s corrective purpose due to their form orientation in this study (see Chapter 6), it was not easy for them to locate the error because the long implicit recast targeted redundant linguistic features in this study. According to VanPatten’s (2004) Preference for Non-redundancy Principle, redundant forms are less likely to be processed due to the learners’ limited processing capacity. The structures the implicit recasts targeted in this study were not frequently attended to. As the exit questionnaire results reveal, although 76% of participants in the IR group claimed that they noticed the teacher correcting grammatical errors, only 21% of participants recalled the correction of verb forms / 3rd person -s, and 9% of participants reported the correction of embedded questions. Arguably, the non-salience of implicit recasts and the redundancy of target structures constrained the usefulness of implicit recasts for implicit knowledge.
It is quite interesting that there was a large effect of oral interaction. The task control group was more accurate than the test control group with a large effect size in OP posttest 1 for both target structures in addition to its improvement over time. The effect of interaction could possibly be explained by the “foreign language” context. This study was conducted in a Chinese university, where English is generally regarded as an object to learn rather than a tool to use. According to Lyster and Mori’s (2006) Counterbalance Hypothesis, instruction that pushes learners towards the direction opposite to that familiar to them triggers L2 learning.

The task-based instruction in this study, different from their daily teaching setting, may have motivated the participants to learn, as shown in a learner’s comment after Session 2, “I like your lecture, because it is different from ours, (and) it is so interesting.” It can be speculated that the learners might have attended to form although they didn’t receive any kind of feedback because they were form-oriented in a foreign language classroom. Whether or not learners spontaneously focus on form during task-based communication in a foreign language context can be investigated in future using a stimulated recall protocol. This finding, in line with Erlam and Loewen (2010), Mackey and Goo (2007) and Chen (2010), provides evidence for the claim that oral interaction without CF is effective in facilitating L2 learning.

The task control group gained significant improvement in the production of embedded questions from EI posttest 1 to posttest 2. According to Lightbown’s (1998) Delayed Effect Hypothesis, lack of immediate repair cannot be taken as evidence that the pedagogical intervention is ineffective because it takes time for its efficacy to become observable. This long-term view of focused tasks sheds light on why the task control group had considerable gains between EI posttests in this study.

Question 1c asks whether there is a difference between the effects of CR and IR on L2 implicit knowledge development. The answer is affirmative. The CR group outperformed the IR group in accuracy of embedded questions in OP posttest 2. The advantage of CR over IR in this study is in contrast to Erlam and Loewen’s (2010) finding that there is no significant effect for the type of recasts. It should be noted that their implicit recasts were operationalized as “recast with interrogative intonation” and the interrogative intonation might have rendered their feedback more salient (R. Ellis, 1995) than those implicit recasts with declarative intonation in this study. The explicitness of their “implicit recasts” may have narrowed the CR - IR difference. This is supported by their evidence that “for noun-adjective agreement, the awareness level was similar for the two treatment groups at roughly 65%” (p.
Besides, the lack of difference between types of recasts could have resulted from their 1-hour short-time treatment. If their treatment had been prolonged, there might have been a significant difference between the CR and IR groups, as shown in this study and Taddarth’s (2010), who argued for the differential effects of CR and IR in terms of uptake and L2 learning.

The relative effects of corrective recasts and implicit recasts on implicit knowledge depend on the target structure. There was no significant difference between the CR and IR groups in acquisition of 3rd person -s, but there was a significant difference in acquisition of embedded questions. Embedded questions are a salient structure. It was likely that the form-oriented EFL learners had no difficulty in attending to the question word and the word order in an embedded question with or without implicit/explicit recasts while performing the focused tasks in this study. Consequently, the CR, IR and task control groups all improved their accuracy in the target structure in OP posttest 1. In OP posttest 2, however, the effect of implicit recasts was not sustainable due to the ambiguity of the CF on a syntactical feature as complex as embedded questions. At least, some of the implicit recast receivers may have become confused about the word order in embedded questions long after completing the focused tasks, as shown in Chapter 6. In contrast, the corrective recast receivers can maintain their development in the complex structure over time due to the transparent corrective function of the CF. As a result, the complexity of embedded questions affected the durability of the effect of implicit recasts, but not that of explicit recasts, rendering the latter superior in OP posttest 2. However, when the target structure was simple (3rd person -s), the advantage of corrective recasts over implicit recasts was not evident in comparison to a large effect of oral interaction. The acquisition of 3rd person -s was facilitated by oral interaction, with or without implicit/explicit recasts.

Unexpectedly, the test control group demonstrated a medium-ES improvement in accurate use of embedded questions from OP posttest 1 to posttest 2 but showed none from EI posttest 1 to posttest 2. One likely explanation is that the test control group prepared for the OP delayed posttest. The OP texts were the only materials that the test control group could learn in the quasi-experiment. They may have intentionally recalled and memorized the text after the immediate posttest because they knew that the final test would be retelling the same story. Contamination of memorizing the OP text is more evident for embedded questions, the acquisition of which lies in word order, and for the test control group, who did not receive
any treatment and so had no significant improvement in the immediate posttest. Their fluency in reproducing the story in the delayed posttest provides evidence for this speculation. Future research needs to counterbalance testing tasks and implement an exit questionnaire on the control group to interpret any unexpected results.

5.4 Conclusion

The learners receiving corrective recasts improved their implicit knowledge significantly more than did the learners receiving no feedback (i.e. the task control group that just completed the communicative tasks and the test control group that took the tests only). The corrective recasts were noticed due to their explicitness. They may have enabled learners to make the cognitive comparison, an important mechanism for language learning (R. Ellis, 1994).

The lack of difference between the IR and task control groups indicates that implicit recasts seemed to play a limited role in the development of implicit knowledge. It was likely that the non-salience of implicit recasts and the redundancy or complexity of the target structure in a communicative setting constrained the effectiveness of implicit recasts.

The CR group outperformed the IR group in the OP delayed posttest for embedded questions. It shows that the explicitness of recasts is important for the acquisition of implicit knowledge.

In the acquisition of a complex syntactic structure (embedded questions), corrective recasts had a superior effect over implicit recasts or tasks alone. In contrast, in the acquisition of a simple morphosyntactic structure (3rd person -s), interaction, interaction plus implicit recasts, and interaction plus corrective recasts were equally effective. The relative effects of explicit and implicit recasts on implicit knowledge development depend on the target structure.

The task control group had higher scores for implicit knowledge than the test control group with a large effect size. The benefit of interaction without any CF may be attributed to the learners’ form orientation in the EFL context. The learners may have spontaneously focused on form when performing the focused tasks.
Chapter 6  Effect of Recasts on Explicit Knowledge: Results and Discussion

Chapter 6 addresses Question 2, which explores the relationship between feedback types and explicit knowledge improvement. The effect of corrective and implicit recasts on explicit knowledge was measured by means of a written production task (WP) and an untimed grammaticality judgment test (UGJ). In both tests, participants were allowed unlimited time to complete the task and had opportunities to deploy their explicit knowledge. As reported in Section 4.5.4 of Chapter 4, the principal-components factor analysis of the participants’ pretest scores for 3rd person -s confirmed that WP and UGJ (ungrammatical item) scores can be represented by one factor, indicating that these two tests are likely to measure explicit knowledge.

RQ2. What effects do corrective recasts and implicit recasts have on learners’ explicit knowledge of L2 grammatical features (3rd person -s and embedded questions)?

2a. What effect do corrective recasts have on learners’ explicit knowledge of L2 grammatical features?

2b. What effect do implicit recasts have on learners’ explicit knowledge of L2 grammatical features?

2c. Is there any difference between the effects of corrective and implicit recasts on learners’ explicit knowledge?

6.1 Statistical treatment

The participants’ WP / UGJ scores for 3rd person -s / embedded questions were analyzed using SPSS Version 21. First of all, a mixed 3 x 4 ANOVA was conducted with “group” as a between-subject variable and “time” as a within-subject variable. When the time x group interaction was significant, the main effect of time was examined in each group by means of separate repeated-measures ANOVAs or non-parametric Friedman tests, and the main effect of group at each testing point was analysed using one-way ANOVAs or non-parametric Kruskal-Wallis H tests. When the test for pretest scores reported a significant between-group difference, posttest scores were compared across groups by means of parametric or non-parametric ANCOVAs with the pretest scores as a covariate. Post-hoc comparisons were carried out to investigate whether corrective and implicit recasts have differential effects on
the learning of the target structures.

For a significant time by group interaction found in mixed ANOVAs, an effect size (partial eta squared $\eta^2_p$) is reported to reveal the magnitude of interaction. A $\eta^2_p$ value of .01 to .06 is evaluated as small ES, .06 to .14 is medium ES, and above .14 is large ES. A $\eta^2_p$ value is also reported for the significant main effect of time/group. For a significant difference between groups or a significant gain from pretest to posttest, Cohen’s $d$ is calculated using the formula of “(Mean 1 – Mean 2) / SD” corrected by pooled standard deviation to examine the magnitude of a between-group difference (Field, 2009), corrected by the correlation coefficient to examine the magnitude of a pretest-posttest difference (Morris & Deshon, 2002), and corrected by the correlation of covariate and dependant variable to examine the magnitude of treatment in the case of ANCOVA (Lipsey & Wilson, 2001). A $d$ value of .2 to .5 is interpreted as small ES, .5 to .8 is medium ES, and any value above .8 is large ES.

6.2 Results

The WP and UGJ (total and ungrammatical item) results for 3rd person -s (the target of Study 1) and embedded questions (the target of Study 2) are reported respectively.

6.2.1 Results for 3rd person -s: Written production

Descriptive statistics

Scores for 3rd person -s in the written production tests are displayed in Table 41. The corrective recast group’s (CR) mean scores rose from 67.64 in the pretest to 85.77 in posttest 1, and then edged down to 83.47 in posttest 2 as illustrated in Figure 7. The implicit recast group’s (IR) scores also improved dramatically from 76.46 in the pretest to 90.51 in posttest 1, and then decreased slightly to 87.56 in posttest 2. The task control group followed the same development pattern with a mean score of 75.91 in the pretest, 89.16 in posttest 1 and 86.91 in posttest 2. The test control group did not change much with a mean score of 85.51 in the pretest, 88.53 in posttest 1 and 85.93 in posttest 2.
Table 41: Descriptive Statistics of WP Scores for 3rd Person -s

<table>
<thead>
<tr>
<th></th>
<th>Pretest</th>
<th>Posttest 1</th>
<th>Posttest 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
</tr>
<tr>
<td>CR group (n=17)</td>
<td>67.64</td>
<td>18.82</td>
<td>85.77</td>
</tr>
<tr>
<td>IR group (n=14)</td>
<td>76.46</td>
<td>13.00</td>
<td>90.51</td>
</tr>
<tr>
<td>TkC group (n=7)</td>
<td>75.91</td>
<td>10.54</td>
<td>89.16</td>
</tr>
<tr>
<td>TtC group (n=8)</td>
<td>85.51</td>
<td>2.30</td>
<td>88.53</td>
</tr>
<tr>
<td>Total (n=46)</td>
<td>74.69</td>
<td>15.20</td>
<td>88.20</td>
</tr>
</tbody>
</table>

Note. CR = corrective recast; IR = implicit recast; TkC = task control; TtC = test control.

Figure 7: Means for 3rd Person -s in WP Tests over Time by Group

Inferential statistics

Omnibus results

A mixed 3 x 4 ANOVA of the scores for 3rd person -s in the WP tests was performed across groups over time. The omnibus analysis showed the time x group interaction was significant,
The significant time x group interaction suggests that it is necessary to examine the effect of time in each group and the effect of group at each testing point.

**Results for the effect of time in each group**

Non-parametric Friedman tests were carried out for the CR, IR and task control groups because the CR and IR groups’ pretest scores and the task control group’s immediate posttest scores did not have a normal distribution. The Friedman test for the CR group found a significant difference between testing points, $\chi^2(2) = 20.63, p = .00$. Wilcoxon signed-ranks or paired-samples $t$ tests as a post-hoc procedure were conducted with a Bonferroni correction applied, which resulted in the alpha level set at $p < .02$. Results revealed that the CR group significantly improved its accuracy in producing 3rd person -s in posttest 1, $Z = 3.62, p = .00, d = 1.35$, and in posttest 2, $Z = 3.24, p = .00, d = 1.53$. The decrease from posttest 1 to posttest 2 was not significant, $t(16) = .76, p > .05, d = .17$, indicating that the effect of corrective recasts was durable.

The Friedman test for the IR group also showed a significant time effect, $\chi^2(2) = 14.51, p = .00$. With the alpha level set at $p < .02$, post-hoc analyses found a significant difference between pretest and posttest 1, $Z = 3.23, p = .00, d = 2.39$, and between pretest and posttest 2, $Z = 2.67, p = .01, d = 1.01$, but not between the posttests, $t(13) = 1.45, p > .05, d = .50$.

The Friedman test for the task control group also reported a significant time effect, $\chi^2(2) = 11.19, p = .00$. With the alpha level set at $p < .02$, post-hoc analyses showed the task control group’s improvement from pretest to posttest 1 was significant, $Z = 2.37, p = .02$, with a large effect size, $d = 1.82$. The improvement from pretest to posttest 2 was significant, $t(6) = 4.18, p = .01, d = 1.13$. There was no significant difference between the posttests, $Z = .74, p > .05, d = .31$.

A repeated-measures ANOVA for the test control group showed no significant time effect, $F(2, 14) = .57, p > .05, \eta^2_p = .08$.

**Results for the effect of group at each testing point**

WP pretest scores were compared across groups using a Kruskal-Wallis $H$ test because the
CR and IR groups’ scores had a skewed distribution. The test revealed a significant between-group difference in the use of 3rd person -s prior to the treatment, $H(3) = 7.71, p = .05$. Post-hoc procedures using Mann-Whitney $U$ tests and an independent-samples $t$ test with a Bonferroni correction ($p < .01$) showed a significant difference between the CR and test control groups, $U = 24.50, p = .01, d = 1.33$, and a marginally significant difference between the IR and test control groups, $U = 27.00, p = .05, d = .97$, and between the control groups, $t(6.50) = 2.36, p = .03, d = 1.26$ (degrees of freedom were adjusted due to the violation of the equal variance assumption). In view of the pre-treatment differences, WP posttest scores were compared across groups by means of ANCOVAs with pretest scores as a covariate. The adjusted means of unstandardized residuals for posttest 1 and of accuracy scores for posttest 2 controlling for the effect of pretest scores are listed in Table 42.

### Table 42: Adjusted Means for 3rd Person -s in WP Posttests

<table>
<thead>
<tr>
<th></th>
<th>Posttest 1</th>
<th>Posttest 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean $^a$</td>
<td>SD</td>
</tr>
<tr>
<td>CR group (n=17)</td>
<td>1.50</td>
<td>12.88</td>
</tr>
<tr>
<td>IR group (n=14)</td>
<td>.93</td>
<td>9.64</td>
</tr>
<tr>
<td>TkC group (n=7)</td>
<td>.83</td>
<td>5.60</td>
</tr>
<tr>
<td>TtC group (n=8)</td>
<td>-5.53</td>
<td>11.20</td>
</tr>
</tbody>
</table>

Note. $^a$ Means of unstandardized residuals were compared in a non-parametric ANCOVA. $^b$ Covariates were evaluated at the value of 74.69. CR = corrective recast; IR = implicit recast; TkC = task control; TtC = test control.

A Quade non-parametric ANCOVA was conducted for WP posttest 1 because the task control group’s scores did not have a normal distribution. The test determined that there was no significant between-group difference in posttest 1, $F(3, 42) = .85, p > .05, \eta^2 = .06$.

A one-way ANCOVA for WP posttest 2 found no significant between-group difference among adjusted means, $F(3, 41) = .61, p > .05, \eta^p^2 = .04$, indicating that the four groups did not differ significantly in posttest 2.
Table 43: A Summary of within- and between-Subject Differences in WP Tests for 3\textsuperscript{rd} Person -s

<table>
<thead>
<tr>
<th></th>
<th>df</th>
<th>p value</th>
<th>Effect size ((d))</th>
<th>Effect size ((\eta^2))</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Within-subject</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CR group: pretest - posttest 1</td>
<td>16</td>
<td>.00***</td>
<td>1.35(^0)</td>
<td></td>
</tr>
<tr>
<td>CR group: pretest - posttest 2</td>
<td>16</td>
<td>.00***</td>
<td>1.53(^0)</td>
<td></td>
</tr>
<tr>
<td>CR group: posttest1 - posttest 2</td>
<td>16</td>
<td>.46</td>
<td>.17</td>
<td></td>
</tr>
<tr>
<td>IR group: pretest - posttest 1</td>
<td>13</td>
<td>.00***</td>
<td>2.39(^0)</td>
<td></td>
</tr>
<tr>
<td>IR group: pretest - posttest 2</td>
<td>13</td>
<td>.01**</td>
<td>1.01(^0)</td>
<td></td>
</tr>
<tr>
<td>IR group: posttest1 - posttest 2</td>
<td>13</td>
<td>.17</td>
<td>.50</td>
<td></td>
</tr>
<tr>
<td>TkC group: pretest - posttest 1</td>
<td>6</td>
<td>.02*</td>
<td>1.82(^0)</td>
<td></td>
</tr>
<tr>
<td>TkC group: pretest - posttest 2</td>
<td>6</td>
<td>.01**</td>
<td>1.13(^0)</td>
<td></td>
</tr>
<tr>
<td>TkC group: posttest1 - posttest 2</td>
<td>6</td>
<td>.46</td>
<td>.31</td>
<td></td>
</tr>
<tr>
<td>TiC group: omnibus</td>
<td>2</td>
<td>.58</td>
<td>.08</td>
<td></td>
</tr>
<tr>
<td><strong>Between-subject</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pretest: CR - IR</td>
<td>29</td>
<td>.22</td>
<td>.55</td>
<td></td>
</tr>
<tr>
<td>Pretest: CR - TkC</td>
<td>22</td>
<td>.49</td>
<td>.54</td>
<td></td>
</tr>
<tr>
<td>Pretest: CR - TiC</td>
<td>23</td>
<td>.01**</td>
<td>1.33(^0)</td>
<td></td>
</tr>
<tr>
<td>Pretest: IR - TkC</td>
<td>19</td>
<td>.65</td>
<td>.12</td>
<td></td>
</tr>
<tr>
<td>Pretest: IR - TiC</td>
<td>20</td>
<td>.05*</td>
<td>.97(^0)</td>
<td></td>
</tr>
<tr>
<td>Pretest: TkC - TiC</td>
<td>13</td>
<td>.03*</td>
<td>1.26(^0)</td>
<td></td>
</tr>
<tr>
<td>Posttest 1: omnibus</td>
<td>3</td>
<td>.47</td>
<td>.06</td>
<td></td>
</tr>
<tr>
<td>Posttest 2: omnibus</td>
<td>3</td>
<td>.62</td>
<td>.04</td>
<td></td>
</tr>
</tbody>
</table>

Note. CR = corrective recast; IR = implicit recast; TkC = task control; TiC = test control.
* \(p \leq .05\).  ** \(p \leq .01\).  *** \(p \leq .001\).  \(^0\) = large effect size.

A summary of significant differences in the written production tests for 3\textsuperscript{rd} person -s

- The CR and IR groups significantly improved their accuracy from pretest to posttests.
- The task control group’s improvement was also significant in posttest 1 and 2.
- The test control group did not change significantly from pretest to posttests.
- There were (marginally) significant between-group differences prior to the treatment.
- The four groups did not differ significantly in posttest 1, nor in posttest 2.
6.2.2 Results for 3\textsuperscript{rd} person -\textit{s}: Untimed grammaticality judgment

\textit{Descriptive statistics}

Table 44 reports total scores for 3\textsuperscript{rd} person -\textit{s} in untimed grammaticality judgment tests. The CR group’s mean scores increased dramatically from 61.11 in the pretest to 84.00 in posttest 1, and then edged down to 77.56 in posttest 2 as illustrated in Figure 8. The IR group’s scores also grew quickly from 72.78 in the pretest to 83.33 in posttest 1, and then decreased slightly to 79.17 in posttest 2. Following the same pattern, the task control group had a mean score of 70.67 in the pretest, and higher means of 75.99 in posttest 1 and 74.00 in posttest 2. The test control group’s means edged down from 77.62 in the pretest to 74.29 in posttest 1 and then down to 71.43 in posttest 2.

Table 44: Descriptive Statistics of UGJ Total Scores for 3\textsuperscript{rd} Person -\textit{s}

<table>
<thead>
<tr>
<th></th>
<th>Pretest</th>
<th></th>
<th>Posttest 1</th>
<th></th>
<th>Posttest 2</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>CR group ((n=15))</td>
<td>61.11</td>
<td>14.84</td>
<td>84.00</td>
<td>8.19</td>
<td>77.56</td>
<td>18.06</td>
</tr>
<tr>
<td>IR group ((n=12))</td>
<td>72.78</td>
<td>13.69</td>
<td>83.33</td>
<td>9.53</td>
<td>79.17</td>
<td>17.00</td>
</tr>
<tr>
<td>TkC group ((n=10))</td>
<td>70.67</td>
<td>7.50</td>
<td>75.99</td>
<td>7.66</td>
<td>74.00</td>
<td>13.50</td>
</tr>
<tr>
<td>TtC group ((n=7))</td>
<td>77.62</td>
<td>7.87</td>
<td>74.29</td>
<td>21.41</td>
<td>71.43</td>
<td>15.50</td>
</tr>
<tr>
<td>Total ((n=44))</td>
<td>69.09</td>
<td>13.36</td>
<td>80.45</td>
<td>11.78</td>
<td>76.21</td>
<td>16.12</td>
</tr>
</tbody>
</table>

\textit{Note.} CR = corrective recast; IR = implicit recast; TkC = task control; TtC = test control.
Inferential statistics

Omnibus results

A mixed ANOVA of the UGJ total scores for 3rd person -s revealed that the interaction between time and group reached significance, $F (6, 80) = 4.21, p = .00, \eta^2_p = .24$. The significant interaction was further analyzed by examining the main effect of time in each group and the main effect of group at each testing point.

Results for the effect of time in each group

A repeated-measures ANOVA of the UGJ total scores was conducted for the CR group. The analysis indicated a significant main effect for time, $F (2, 28) = 21.54, p = .00, \eta^2_p = .61$. Post-hoc analyses showed the CR group’s accuracy in judging the grammatical rule of 3rd person -s significantly improved from pretest to posttest 1, $p = .00, d = 2.80$, and from pretest to posttest 2, $p = .01, d = .91$. There was no significant difference between the posttests,
A repeated-measures ANOVA for the IR group also found a significant effect for time, $F(2, 22) = 5.38, p = .01, \eta_p^2 = .33$. Post-hoc comparisons revealed the difference between pretest and posttest 1 was significant, $p = .03, d = 1.11$, but the difference between pretest and posttest 2 was not, $p > .05, d = .38$. No significant difference was observed between the posttests either, $p > .05, d = .48$.

A repeated-measures ANOVA was carried out for the task control group with degrees of freedom corrected using Greenhouse-Geisser estimates of sphericity ($\varepsilon = .66$) because the assumption of sphericity had been violated as determined by the Mauchly’s test, $\chi^2(2) = 5.84, p = .05$. The analysis for the task control group showed no significant difference over time, $F(1.32, 11.86) = .76, p > .05, \eta_p^2 = .08$.

A repeated-measures ANOVA for the test control group revealed no significant difference between testing times, $F(2, 40) = .10, p > .05, \eta_p^2 = .01$.

**Results for the effect of group at each testing point**

The participants’ scores for 3rd person -s in the UGJ pretest were compared across groups by means of a Kruskal-Wallis $H$ test as the test control group’s scores did not have a normal distribution. The analysis determined that there was a significant effect of group, $H(3) = 9.56, p = .02$. Post-hoc comparisons were conducted using Mann-Whitney $U$ tests and independent-samples $t$ tests with the alpha level set at $p < .01$. It was found that the CR group differed from the IR group, $t(25) = 2.10, p = .05, d = .82$, from the task control group, $t(21.81) = 2.12, p = .05, d = .81$ (degrees of freedom were adjusted due to the violation of equality of variances), and from the test control group, $U = 21.00, p = .02, d = 1.39$, of marginal significance. The difference between control groups was not significant, $U = 16.50, p = .07$, but the effect size was fairly large, $d = .90$. Consequently, UGJ posttest scores were analyzed across groups by means of ANCOVAs controlling for the effect of pretest scores.

A preliminary analysis of the scores in UGJ posttest 1 revealed that the homogeneity-of-slopes assumption was violated as determined by the significant interaction between the independent variable and the covariate, $F(3, 36) = 4.23, p = .01$. The heterogeneous slopes
implied that the mean differences between groups varied as a function of covariate scores. As a result, follow-up tests were conducted to assess between-group differences for particular scores on the covariate: Mean - 1SD (55.73), Mean (69.09), and Mean + 1SD (82.45). Accordingly, a $p$ value of .02 was set as significance.

When the pretest score was 55.73, estimated means 81.90 for the CR group, 77.18 for the IR group, 81.43 for the task control group and 33.41 for the test control group in UGJ posttest 1 (see Table 45) differed significantly, $F(3, 36) = 6.11, p = .00, \eta^2_p = .34$. Pairwise comparisons indicated significant differences between the CR and test control groups, $p = .00, d = 4.91$, between the IR and test control groups, $p = .00, d = 4.40$, and between the task control and test control groups, $p = .00, d = 4.81$.

When the pretest score was 69.09, estimated means 87.11 for the CR group, 82.00 for the IR group, 76.56 for the task control group and 58.36 for the test control group in posttest 1 differed significantly, $F(3, 36) = 8.15, p = .00, \eta^2_p = .40$. Pairwise comparisons disclosed significant differences between the CR and task control groups, $p = .01, d = 1.07$, between the CR and test control groups, $p = .00, d = 2.91$, between the IR and test control groups, $p = .00, d = 2.38$, and between the task control and test control groups, $p = .01, d = 1.82$.

When the pretest score was 82.45, estimated means 92.31 for the CR group, 86.82 for the IR group, 71.70 for the task control group and 83.31 for the test control group in posttest 1 marginally significantly differed, $F(3, 36) = 2.93, p = .05$, with a large effect size, $\eta^2_p = .20$. Pairwise comparisons displayed a significant difference between the CR and task control groups, $p = .01, d = 2.09$, and a marginally significant difference between the IR and task control groups, $p = .03, d = 1.53$. There was also a large effect size for the CR - test control difference, $d = .91$. The task control group had a lower mean score than the test control group with a large effect size, $d = 1.16$. 

142
### Table 45: Adjusted Means for 3rd Person -s in UGJ Posttests

<table>
<thead>
<tr>
<th>Group</th>
<th>Mean^a</th>
<th>Std Error</th>
<th>Mean^b</th>
<th>Std Error</th>
<th>Mean^c</th>
<th>Std Error</th>
<th>Mean^b</th>
<th>Std Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>CR group</td>
<td>81.90</td>
<td>2.57</td>
<td>87.11</td>
<td>2.75</td>
<td>92.31</td>
<td>4.31</td>
<td>82.83</td>
<td>4.03</td>
</tr>
<tr>
<td>(n=15)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IR group</td>
<td>77.18</td>
<td>4.41</td>
<td>82.00</td>
<td>2.79</td>
<td>86.82</td>
<td>3.34</td>
<td>76.73</td>
<td>4.24</td>
</tr>
<tr>
<td>(n=12)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TkC group</td>
<td>81.43</td>
<td>6.84</td>
<td>76.56</td>
<td>3.01</td>
<td>71.70</td>
<td>5.69</td>
<td>72.96</td>
<td>4.60</td>
</tr>
<tr>
<td>(n=10)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TtC group</td>
<td>33.41</td>
<td>11.14</td>
<td>58.36</td>
<td>5.42</td>
<td>83.31</td>
<td>4.22</td>
<td>65.80</td>
<td>5.71</td>
</tr>
<tr>
<td>(n=7)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. CR = corrective recast; IR = implicit recast; TkC = task control; TtC = test control.

a. Covariates were evaluated at the value: UGJ pretest scores = 55.73.

b. Covariates were evaluated at the value: UGJ pretest scores = 69.09.

c. Covariates were evaluated at the value: UGJ pretest scores = 82.45.

An ANCOVA of UGJ total scores in posttest 2 controlling for the effect of pretest scores determined there was no significant difference among adjusted means for the four groups, $F(3, 39) = 1.99$, $p > .05$, although the effect size of between-group differences was medium, $\eta_p^2 = .13$. 

---

143
<table>
<thead>
<tr>
<th></th>
<th>df</th>
<th>p value</th>
<th>Effect size (d)</th>
<th>Effect size (η²)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Within-subject</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CR group: pretest - posttest 1</td>
<td>14</td>
<td>.00***</td>
<td>2.80*</td>
<td></td>
</tr>
<tr>
<td>CR group: pretest - posttest 2</td>
<td>14</td>
<td>.01**</td>
<td>.91#</td>
<td></td>
</tr>
<tr>
<td>CR group: posttest1 - posttest 2</td>
<td>14</td>
<td>.29</td>
<td>.79</td>
<td></td>
</tr>
<tr>
<td>IR group: pretest - posttest 1</td>
<td>11</td>
<td>.03*</td>
<td>1.11#</td>
<td></td>
</tr>
<tr>
<td>IR group: pretest - posttest 2</td>
<td>11</td>
<td>.21</td>
<td>.38</td>
<td></td>
</tr>
<tr>
<td>IR group: posttest1 - posttest 2</td>
<td>11</td>
<td>.62</td>
<td>.48</td>
<td></td>
</tr>
<tr>
<td>TkC group: omnibus</td>
<td>1.3</td>
<td>.44</td>
<td>.08</td>
<td></td>
</tr>
<tr>
<td>TiC group: omnibus</td>
<td>2</td>
<td>.32</td>
<td>.01</td>
<td></td>
</tr>
<tr>
<td><strong>Between-subject</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pretest: CR - IR</td>
<td>25</td>
<td>.05*</td>
<td>.82#</td>
<td></td>
</tr>
<tr>
<td>Pretest: CR - TkC</td>
<td>21.8</td>
<td>.05*</td>
<td>.81#</td>
<td></td>
</tr>
<tr>
<td>Pretest: CR - TiC</td>
<td>20</td>
<td>.02*</td>
<td>1.39#</td>
<td></td>
</tr>
<tr>
<td>Pretest: IR - TkC</td>
<td>17.6</td>
<td>.65</td>
<td>.19</td>
<td></td>
</tr>
<tr>
<td>Pretest: IR - TiC</td>
<td>17</td>
<td>.46</td>
<td>.43</td>
<td></td>
</tr>
<tr>
<td>Pretest: TkC - TiC</td>
<td>15</td>
<td>.07</td>
<td>.90#</td>
<td></td>
</tr>
<tr>
<td>Posttest 1: CR - IR</td>
<td>25</td>
<td>.36a</td>
<td>.48</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>.20b</td>
<td>.52</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>.32c</td>
<td>.56</td>
<td></td>
</tr>
<tr>
<td>Posttest 1: CR - TkC</td>
<td>23</td>
<td>.95</td>
<td>.05</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>.01**</td>
<td>1.07#</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>.01**</td>
<td>2.09#</td>
<td></td>
</tr>
<tr>
<td>Posttest 1: CR - TiC</td>
<td>20</td>
<td>.00***</td>
<td>4.91#</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>.00***</td>
<td>2.91#</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>.14</td>
<td>.91#</td>
<td></td>
</tr>
<tr>
<td>Posttest 1: IR - TkC</td>
<td>20</td>
<td>.61</td>
<td>.43</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>.19</td>
<td>.55</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>.03*</td>
<td>1.53#</td>
<td></td>
</tr>
<tr>
<td>Posttest 1: IR - TiC</td>
<td>17</td>
<td>.00***</td>
<td>4.40#</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>.00***</td>
<td>2.38#</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>.52</td>
<td>.35</td>
<td></td>
</tr>
<tr>
<td>Posttest 1: TkC - TiC</td>
<td>15</td>
<td>.00***</td>
<td>4.81#</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>.01**</td>
<td>1.82#</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>.11</td>
<td>1.16#</td>
<td></td>
</tr>
<tr>
<td>Posttest 2: omnibus</td>
<td>3</td>
<td>.13</td>
<td>.13</td>
<td></td>
</tr>
</tbody>
</table>

Note. CR = corrective recast; IR = implicit recast; TkC = task control; TiC = test control.
*p ≤ .05.  ** p ≤ .01.   *** p ≤ .001.  # = large effect size.
a. Covariates were evaluated at the value: UGJ pretest scores = 55.73.
b. Covariates were evaluated at the value: UGJ pretest scores = 69.09.
c. Covariates were evaluated at the value: UGJ pretest scores = 82.45.
A summary of significant differences in the UGJ total scores for 3rd person -s

- The CR group’s scores in the UGJ test for 3rd person -s significantly improved from pretest to posttests.
- The IR group’s scores for 3rd person -s in UGJ posttest 1 were significantly higher than in the pretest but the scores in UGJ posttest 2 were not.
- The task control group’s improvement from pretest to posttests was not significant.
- The test control group did not vary significantly from pretest to posttests.
- There were marginally significant between-group differences in the ability to judge the rule of 3rd person -s prior to the treatment.
- The CR, IR and task control groups all significantly outperformed the test control group in posttest 1 when the covariate was set at the value of M - 1 SD.
- The CR group outperformed both control groups significantly, but the IR and task control groups significantly outperformed the test control group only in posttest 1 when the covariate was set at the Mean.
- The CR group outperformed both of the control groups significantly or with a large effect size, but the IR group marginally significantly outperformed the task control group only when the covariate was set at the value of M + 1 SD.
- The four groups’ adjusted means in UGJ posttest 2 did not differ significantly.

6.2.3 Results for 3rd person -s: Untimed grammaticality judgment (ungrammatical items)

Descriptive statistics

Table 47 reports ungrammatical item scores for 3rd person -s in the untimed grammaticality judgment tests. The CR group’s mean scores increased dramatically from 56.08 in the pretest to 76.45 in posttest 1, and then edged down to 75.12 in posttest 2 as illustrated in Figure 9. The IR group’s scores also grew quickly from 67.98 in the pretest to 75.92 in posttest 1, but then decreased to 71.15 in posttest 2. Following the same pattern, the task control group had a
mean score of 65.56 in the pretest, and higher means of 74.45 in posttest 1 and 69.63 in posttest 2. The test control group’s means edged down from 72.21 in the pretest to 67.28 in posttest 1 and then down to 61.71 in posttest 2.

Table 47: Descriptive Statistics of UGJ (Ungrammatical Item) Scores for 3rd Person -s

<table>
<thead>
<tr>
<th></th>
<th>Pretest</th>
<th></th>
<th>Posttest 1</th>
<th></th>
<th>Posttest 2</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>CR group (n= 21)</td>
<td>56.08</td>
<td>20.03</td>
<td>76.45</td>
<td>23.35</td>
<td>75.12</td>
<td>27.93</td>
</tr>
<tr>
<td>IR group (n= 21)</td>
<td>67.98</td>
<td>19.79</td>
<td>75.92</td>
<td>27.62</td>
<td>71.15</td>
<td>28.36</td>
</tr>
<tr>
<td>TkC group (n= 15)</td>
<td>65.56</td>
<td>13.30</td>
<td>74.45</td>
<td>19.78</td>
<td>69.63</td>
<td>23.28</td>
</tr>
<tr>
<td>TtC group (n= 9)</td>
<td>72.21</td>
<td>14.41</td>
<td>67.28</td>
<td>30.98</td>
<td>61.71</td>
<td>25.06</td>
</tr>
<tr>
<td>Total (n= 66)</td>
<td>64.22</td>
<td>18.52</td>
<td>74.58</td>
<td>24.78</td>
<td>70.78</td>
<td>26.44</td>
</tr>
</tbody>
</table>

Note. CR = corrective recast; IR = implicit recast; TkC = task control; TtC = test control.

Figure 9: Means for 3rd Person -s in UGJ (Ungrammatical Item) Tests over Time by Group
**Inferential statistics**

**Omnibus results**

A mixed ANOVA of the UGJ ungrammatical item scores for 3rd person -s revealed that the interaction between time and group reached significance, $F(6, 124) = 2.71, p = .02, \eta^2_p = .12$. The significant interaction was analyzed by examining the main effect of time in each group and the main effect of group at each testing point.

**Results for the effect of time in each group**

A repeated-measures ANOVA of the UGJ ungrammatical item scores was conducted for the task control group but non-parametric Friedman tests were carried out for the CR, IR and test control groups due to the skewed distribution of their UGJ scores. The Friedman test for the CR group found a significant difference between testing points, $\chi^2(2) = 14.33, p = .00$. Wilcoxon signed-ranks tests as a post-hoc procedure were conducted with a Bonferroni correction applied, which resulted in the alpha level set at $p < .02$. Results revealed that the CR group significantly improved their accuracy in judging the grammatical rule of 3rd person -s from pretest to posttest 1, $Z = 3.48, p = .00, d = 1.06$, and from pretest to posttest 2, $Z = 2.90, p = .00, d = .83$. There was no significant difference between the posttests, $Z = .37, p > .05, d = .08$.

The Friedman test for the IR group did not report a significant effect for time, $\chi^2(2) = 5.06, p > .05$. The repeated-measures ANOVA for the task control group ($F(2, 28) = 1.03, p > .05, \eta^2_p = .07$) and Friedman test for the test control group ($\chi^2(2) = 2.40, p > .05$) showed no significant difference over time either.

**Results for the effect of group at each testing point**

The participants’ ungrammatical item scores for 3rd person -s in the UGJ pretest were compared across groups by means of a Kruskal-Wallis $H$ test as the IR group’s scores did not have a normal distribution. The analysis determined that the between-group difference approached a significant level, $H(3) = 6.67, p = .08$. Post-hoc comparisons were conducted using Mann-Whitney $U$ tests and independent-samples $t$ tests with the alpha level set at $p < .01$. It was found that the CR group had a marginally significantly lower pretest score than the test control group, $t(28) = 2.18, p = .04, d = .92$, and from the IR group, $U = 1.98, p$
Consequently, UGJ posttest scores were analyzed across groups by means of ANCOVAs controlling for the potential effect of pretest scores.

A preliminary analysis of the ungrammatical item scores in UGJ posttest 1 revealed that the homogeneity-of-slopes assumption was violated as determined by the significant interaction between independent variable and covariate, $F(4, 61) = 6.54, p = .00$. The heterogeneous slopes implied that the mean differences between groups varied as a function of covariate scores. As a result, follow-up tests were conducted to assess between-group differences for particular scores on the covariate: Mean - 1SD (45.70), Mean (64.22), and Mean + 1SD (82.70). Accordingly, a $p$ value of .02 was set as significance.

When the pretest score was 45.70, estimated means 69.13 for the CR group, 55.86 for the IR group, 71.20 for the task control group and 29.43 for the test control group in UGJ posttest 1 (see Table 48) differed marginally significantly, $F(3, 58) = 2.64, p = .06, \eta^2_p = .12$. Pairwise comparisons indicated (marginally) significant differences between the CR and test control groups, $p = .02, d = 1.61$, and between the task control and test control groups, $p = .03, d = 1.69$. The IR and test control difference had a large effect size ($d = 1.07$).

When the pretest score was 64.22, estimated means 82.20 for the CR group, 72.54 for the IR group, 74.23 for the task control group and 55.87 for the test control group in posttest 1 differed of marginal significance ($F(3, 58) = 2.61, p = .06, \eta^2_p = .12$). Pairwise comparisons revealed a significant difference between the CR and test control groups, $p = .01, d = 1.07$, but other between-group differences were not significant.

When the pretest score was 82.70, estimated means 95.27 for the CR group, 89.22 for the IR group, 77.25 for the task control group and 82.30 for the test control group in posttest 1 did not significantly differ ($F(3, 58) = .91, p > .05, \eta^2_p = .05$).
A preliminary analysis of the ungrammatical item scores in UGJ posttest 2 also revealed a significant interaction between independent variable and covariate, $F(4, 61) = 7.37, p = .00$. Therefore, between-group differences were assessed for three levels of the covariate: Mean - 1SD (45.70), Mean (64.22), and Mean + 1SD (82.70). Accordingly, a $p$ value of .02 was set as significant.

When the pretest score was 45.70, estimated means 67.30 for the CR group, 49.12 for the IR group, 61.12 for the task control group and 26.81 for the test control group in UGJ posttest 2 (see Table 49) differed marginally significantly, $F(3, 58) = 2.67, p = .06, \eta^2_p = .12$. Pairwise comparisons indicated that the CR group (marginally) significantly outperformed the test control group, $p = .02, d = 1.55$, and the IR group, $p = .05, d = .70$. The IR ($d = .86$) and task control ($d = 1.31$) groups’ adjusted means were also higher than the test control group’s with a large effect size.

### Table 48: Adjusted Means for 3rd Person -s in UGJ Posttest 1 (Ungrammatical Item)

<table>
<thead>
<tr>
<th>Group</th>
<th>Meana</th>
<th>Std Error</th>
<th>Meanb</th>
<th>Std Error</th>
<th>Meanc</th>
<th>Std Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>CR group (n=21)</td>
<td>69.13</td>
<td>5.18</td>
<td>82.20</td>
<td>4.96</td>
<td>95.27</td>
<td>7.74</td>
</tr>
<tr>
<td>IR group (n=21)</td>
<td>55.86</td>
<td>6.99</td>
<td>72.54</td>
<td>4.66</td>
<td>89.22</td>
<td>5.76</td>
</tr>
<tr>
<td>TkC group (n=15)</td>
<td>71.20</td>
<td>9.97</td>
<td>74.23</td>
<td>5.45</td>
<td>77.25</td>
<td>9.04</td>
</tr>
<tr>
<td>TtC group (n=9)</td>
<td>29.43</td>
<td>15.33</td>
<td>55.87</td>
<td>8.11</td>
<td>82.30</td>
<td>8.84</td>
</tr>
</tbody>
</table>

Note. CR = corrective recast; IR = implicit recast; TkC = task control; TtC = test control.

a. Covariates were evaluated at the value: UGJ pretest scores = 45.70.
b. Covariates were evaluated at the value: UGJ pretest scores = 64.22.
c. Covariates were evaluated at the value: UGJ pretest scores = 82.70.

### Table 49: Adjusted Means for 3rd Person -s in UGJ Posttest 2 (Ungrammatical Item)

<table>
<thead>
<tr>
<th>Group</th>
<th>Meana</th>
<th>Std Error</th>
<th>Meanb</th>
<th>Std Error</th>
<th>Meanc</th>
<th>Std Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>CR group (n=21)</td>
<td>67.30</td>
<td>5.50</td>
<td>81.26</td>
<td>5.26</td>
<td>95.22</td>
<td>8.21</td>
</tr>
<tr>
<td>IR group (n=21)</td>
<td>49.12</td>
<td>7.41</td>
<td>67.43</td>
<td>4.95</td>
<td>85.75</td>
<td>6.11</td>
</tr>
<tr>
<td>TkC group (n=15)</td>
<td>61.12</td>
<td>10.58</td>
<td>69.05</td>
<td>5.78</td>
<td>76.98</td>
<td>9.59</td>
</tr>
<tr>
<td>TtC group (n=9)</td>
<td>26.81</td>
<td>16.26</td>
<td>51.19</td>
<td>8.61</td>
<td>75.57</td>
<td>9.38</td>
</tr>
</tbody>
</table>

Note. CR = corrective recast; IR = implicit recast; TkC = task control; TtC = test control.

a. Covariates were evaluated at the value: UGJ pretest scores = 45.70.
b. Covariates were evaluated at the value: UGJ pretest scores = 64.22.
c. Covariates were evaluated at the value: UGJ pretest scores = 82.70.
When the pretest score was 64.22, estimated means 81.26 for the CR group, 67.43 for the IR group, 69.05 for the task control group and 51.19 for the test control group in posttest 2 differed of marginal significance ($F (3, 58) = 3.22, p = .03, \eta^2_p = .14$). Pairwise comparisons displayed a significant difference between the CR and test control groups, $p = .00, d = 1.15$, but other between-group differences were not significant.

When the pretest score was 82.70, estimated means 95.22 for the CR group, 85.75 for the IR group, 76.98 for the task control group and 75.57 for the test control group in posttest 2 did not significantly differ ($F (3, 58) = 1.09, p > .05, \eta^2_p = .05$).
Table 50: A Summary of within- and between-Subject Differences in UGJ (Ungrammatical Item) Tests for 3rd Person -s

<table>
<thead>
<tr>
<th></th>
<th>df</th>
<th>p value</th>
<th>Effect size (d)</th>
<th>Effect size ($\eta^2$)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Within-subject</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CR group: pretest - posttest 1</td>
<td>20</td>
<td>.00***</td>
<td>1.06*</td>
<td></td>
</tr>
<tr>
<td>CR group: pretest - posttest 2</td>
<td>20</td>
<td>.00***</td>
<td>.83*</td>
<td></td>
</tr>
<tr>
<td>CR group: posttest1 - posttest 2</td>
<td>20</td>
<td>.72</td>
<td>.08</td>
<td></td>
</tr>
<tr>
<td>IR group: omnibus</td>
<td>2</td>
<td>.08</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TkC group: omnibus</td>
<td>2</td>
<td>.37</td>
<td>.07</td>
<td></td>
</tr>
<tr>
<td>TtC group: omnibus</td>
<td>2</td>
<td>.30</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pretest: CR - IR</td>
<td>40</td>
<td>.05*</td>
<td>.60</td>
<td></td>
</tr>
<tr>
<td>Pretest: CR - TkC</td>
<td>33.9</td>
<td>.10</td>
<td>.56</td>
<td></td>
</tr>
<tr>
<td>Pretest: CR - TtC</td>
<td>28</td>
<td>.04*</td>
<td>.92*</td>
<td></td>
</tr>
<tr>
<td>Pretest: IR - TkC</td>
<td>34</td>
<td>.41</td>
<td>.14</td>
<td></td>
</tr>
<tr>
<td>Pretest: IR - TtC</td>
<td>28</td>
<td>.70</td>
<td>.24</td>
<td></td>
</tr>
<tr>
<td>Pretest: TkC - TtC</td>
<td>22</td>
<td>.26</td>
<td>.48</td>
<td></td>
</tr>
<tr>
<td>Posttest 1: CR - IR</td>
<td>40</td>
<td>.13a</td>
<td>.54</td>
<td></td>
</tr>
<tr>
<td>Posttest 1: CR - TkC</td>
<td>34</td>
<td>.85</td>
<td>.08</td>
<td></td>
</tr>
<tr>
<td>Posttest 1: CR - TtC</td>
<td>28</td>
<td>.02*</td>
<td>1.61*</td>
<td></td>
</tr>
<tr>
<td>Posttest 1: IR - TkC</td>
<td>34</td>
<td>.21</td>
<td>.63</td>
<td></td>
</tr>
<tr>
<td>Posttest 1: IR - TtC</td>
<td>28</td>
<td>.12</td>
<td>1.07*</td>
<td></td>
</tr>
<tr>
<td>Posttest 2: CR - IR</td>
<td>40</td>
<td>.05**</td>
<td>.70</td>
<td></td>
</tr>
<tr>
<td>Posttest 2: CR - TkC</td>
<td>34</td>
<td>.61</td>
<td>.24</td>
<td></td>
</tr>
<tr>
<td>Posttest 2: CR - TtC</td>
<td>28</td>
<td>.02*</td>
<td>1.55*</td>
<td></td>
</tr>
<tr>
<td>Posttest 2: IR - TkC</td>
<td>34</td>
<td>.36</td>
<td>.46</td>
<td></td>
</tr>
<tr>
<td>Posttest 2: IR - TtC</td>
<td>28</td>
<td>.22</td>
<td>1.31*</td>
<td></td>
</tr>
<tr>
<td>Posttest 2: TkC - TtC</td>
<td>22</td>
<td>.08</td>
<td>.09</td>
<td></td>
</tr>
</tbody>
</table>

**Note.** CR = corrective recast; IR = implicit recast; TkC = task control; TtC = test control.
* $p \leq .05$. ** $p \leq .01$. *** $p \leq .001$. $\# = $ large effect size.

a. Covariates were evaluated at: UGJ pretest scores=45.70. b. Covariates were evaluated at: UGJ pretest scores=64.22.
A summary of significant differences in the UGJ ungrammatical item scores for 3rd person -s

- The CR group’s ungrammatical item scores in the UGJ test for 3rd person -s significantly improved from pretest to posttests.
- The IR, task control, and test control groups’ improvement in judging the grammatical rule of 3rd person -s was not significant.
- The CR group had marginally significantly lower pretest scores for ungrammatical items in the UGJ test than the IR and test control groups.
- When the covariate was set at the value of M - 1 SD, the CR group (marginally) significantly outperformed the test control group in UGJ posttest 1 and both the IR and test control groups in posttest 2; the IR group performed better than the test control group with a large effect size in both posttests; and the difference between the two control groups also had a large effect size in both posttests.
- When the covariate was set the Mean, the CR group’s scores were higher than the test control group’s in UGJ posttest 1 and posttest 2, but the other groups’ were not.
- There was no significant between-group difference in UGJ posttest 1 and posttest 2 when the covariate was set at the value of M + 1 SD.

6.2.4 Results for embedded questions: Written production

Descriptive statistics

Table 51 summarizes group means and standard deviations (SD) in the written production tests for embedded questions. The CR group maintained a sustainable improvement from 72.22 in the pretest to 80.52 in posttest 1 and then to 85.70 in posttest 2 as illustrated in Figure 10. The IR group’s scores increased from 78.00 in the pretest to 87.39 in posttest 1 and then dropped slightly to 85.21 in posttest 2. The task control group crept up from 74.78 in the pretest to 76.36 in posttest 1, and then rose to 89.57 in posttest 2. In contrast, the test control group’s scores edged up from 79.18 in the pretest to 80.29 in posttest 1 and then slipped back.
to 78.64 in posttest 2.

Table 51: Descriptive Statistics of WP Scores for Embedded Questions

<table>
<thead>
<tr>
<th></th>
<th>Pretest</th>
<th></th>
<th>Posttest 1</th>
<th></th>
<th>Posttest 2</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>CR group (n=10)</td>
<td>72.22</td>
<td>12.88</td>
<td>80.52</td>
<td>14.48</td>
<td>85.70</td>
<td>10.56</td>
</tr>
<tr>
<td>IR group (n=11)</td>
<td>78.00</td>
<td>7.86</td>
<td>87.39</td>
<td>8.35</td>
<td>85.21</td>
<td>12.43</td>
</tr>
<tr>
<td>TkC group (n=10)</td>
<td>74.78</td>
<td>14.64</td>
<td>76.36</td>
<td>12.39</td>
<td>89.57</td>
<td>12.38</td>
</tr>
<tr>
<td>TtC group (n=9)</td>
<td>79.18</td>
<td>8.63</td>
<td>80.29</td>
<td>10.04</td>
<td>78.64</td>
<td>8.97</td>
</tr>
<tr>
<td>Total (n=40)</td>
<td>76.02</td>
<td>11.24</td>
<td>81.32</td>
<td>11.81</td>
<td>84.95</td>
<td>11.49</td>
</tr>
</tbody>
</table>

Note. CR = corrective recast; IR = implicit recast; TkC = task control; TtC = test control.

Figure 10: Means for Embedded Questions in WP Tests over Time by Group

Inferential statistics

Omnibus results
A mixed 3 x 4 ANOVA of the scores for embedded questions in the written production tests found a marginally significant effect for the time x group interaction, $F(6, 72) = 2.15, p = .06, \eta^2_p = .15$. Further analyses were performed to assess the effect of time in each group and the effect of group at each testing point.

**Results for the effect of time in each group**

A repeated-measures ANOVA of the CR group’s scores for embedded questions in the written production tests determined that there was a significant main effect of time, $F(2, 18) = 3.47, p = .05, \eta^2_p = .28$. Bonferroni post-hoc analyses indicated the CR group’s scores in posttest 2 were marginally significantly higher than in the pretest, $p = .07$, with a considerably large effect size, $d = .95$. Other comparisons between testing points for the CR group were not significant.

A repeated-measures ANOVA for the IR group found no significant main effect for time, $F(2, 20) = 2.19, p > .05$, but a fairly large effect size, $\eta^2_p = .18$. The large effect size suggests that the difference over time would have been significant with a larger sample size.

A Friedman test was conducted for the task control group because its scores in the pretest and posttest 1 did not have a normal distribution. The test reported a significant time effect, $\chi^2(2) = 14.97, p = .00$. With the alpha level set at $p < .02$, Wilcoxon signed-ranks tests as a post-hoc procedure showed no significant improvement from pretest to posttest 1, but a significant improvement between the posttests, $Z = 2.67, p = .01, d = 1.50$. The difference between pretest and posttest 2 was also significant, $Z = 2.81, p = .01, d = 1.22$.

A Friedman test for the test control group, whose scores in posttest 1 did not have a normal distribution, revealed no significant time effect, $\chi^2(2) = .41, p > .05$, suggesting that the test control group’s scores did not vary significantly from pretest to posttests.

**Results for the effect of group at each testing point**

WP pretest scores were analyzed across groups using a Kruskal-Wallis $H$ test as the task control group’s scores had a skewed distribution. The test determined no significant between-group difference in the use of embedded questions prior to the treatment, $H(3) = 1.45, p > .05$.

A Kruskal-Wallis $H$ test was also conducted for WP posttest 1 because the task control and
test control groups’ scores did not have a normal distribution. The test found no significant between-group difference in posttest 1, $H (3) = 5.64, p > .05$.

A one-way ANOVA for WP posttest 2 found no significant between-group difference, $F (3, 36) = 1.52, p > .05, \eta^2 = .11$.

### Table 52: A Summary of within- and between-Subject Differences in WP Tests for Embedded Questions

<table>
<thead>
<tr>
<th></th>
<th>df</th>
<th>p value</th>
<th>Effect size ($d$)</th>
<th>Effect size ($\eta^2$)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Within-subject</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CR group: pretest - posttest 1</td>
<td>9</td>
<td>.71</td>
<td>.41</td>
<td></td>
</tr>
<tr>
<td>CR group: pretest - posttest 2</td>
<td>9</td>
<td>.07</td>
<td>.95†</td>
<td></td>
</tr>
<tr>
<td>CR group: posttest1 - posttest 2</td>
<td>9</td>
<td>.58</td>
<td>.62</td>
<td></td>
</tr>
<tr>
<td>IR group: omnibus</td>
<td>2</td>
<td>.14</td>
<td></td>
<td>.18†</td>
</tr>
<tr>
<td>TkC group: pretest - posttest 1</td>
<td>9</td>
<td>.89</td>
<td>.11</td>
<td></td>
</tr>
<tr>
<td>TkC group: pretest - posttest 2</td>
<td>9</td>
<td>.01**</td>
<td>1.22#</td>
<td></td>
</tr>
<tr>
<td>TkC group: posttest1 - posttest 2</td>
<td>9</td>
<td>.01**</td>
<td>1.50#</td>
<td></td>
</tr>
<tr>
<td>TiC group: omnibus</td>
<td>2</td>
<td>.81</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Between-subject</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pretest: omnibus</td>
<td>3</td>
<td>.69</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Posttest 1: omnibus</td>
<td>3</td>
<td>.13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Posttest 2: omnibus</td>
<td>3</td>
<td>.23</td>
<td>.11</td>
<td></td>
</tr>
</tbody>
</table>

Note. CR = corrective recast; IR = implicit recast; TkC = task control; TiC = test control.  
* $p \leq .05$.  ** $p \leq .01$.  *** $p \leq .001$.  † = large effect size.

A summary of significant differences in the WP tests for embedded questions

- The CR group improved its accuracy in the written production of embedded questions in posttest 2 with a large effect size.
- The IR group’s improvement in producing embedded questions over time did not reach a significant level, but the effect size was fairly large.
- The task control group improved its accuracy in the use of embedded questions significantly in posttest 2.
- The test control group did not change significantly from pretest to posttests.
- There was no significant between-group difference in the WP pretest for embedded
questions.

- The four groups did not differ significantly in WP posttest 1 and posttest 2 for embedded questions.

6.2.5 Results for embedded questions: Untimed grammaticality judgment

Descriptive statistics

Table 53 presents each group’s means and SDs in the untimed grammaticality judgment tests for embedded questions. The CR group had a mean score of 62.51 in the pretest and obtained a higher mean of 70.84 in posttest 1 and an even higher mean of 75.34 in posttest 2 (see Figure 11). The IR group was stable from pretest (M = 68.94) to posttest 1 (M = 68.54) but had a sudden growth from posttest 1 to posttest 2 (M = 81.06). The task control group’s means fluctuated from 71.33 in the pretest down to 67.55 in posttest 1, and subsequently up to 69.77 in posttest 2. The test control group’s scores crept up from 59.79 in the pretest to 61.68 in posttest 1 and then to 64.16 in posttest 2.

Table 53: Descriptive Statistics of UGJ Scores for Embedded Questions

<table>
<thead>
<tr>
<th></th>
<th>Pretest</th>
<th>Posttest 1</th>
<th>Posttest 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
</tr>
<tr>
<td>CR group (n=20)</td>
<td>62.51</td>
<td>11.79</td>
<td>70.84</td>
</tr>
<tr>
<td>IR group (n=25)</td>
<td>68.94</td>
<td>12.72</td>
<td>68.54</td>
</tr>
<tr>
<td>TkC group (n=15)</td>
<td>71.33</td>
<td>10.90</td>
<td>67.55</td>
</tr>
<tr>
<td>TtC group (n=16)</td>
<td>59.79</td>
<td>13.30</td>
<td>61.68</td>
</tr>
<tr>
<td>Total (n=76)</td>
<td>65.79</td>
<td>12.83</td>
<td>67.50</td>
</tr>
</tbody>
</table>

Note. CR = corrective recast; IR = implicit recast; TkC = task control; TtC = test control.
Inferential statistics

Omnibus results

A mixed ANOVA of the scores for embedded questions in the UGJ tests determined there was a significant effect for the time x group interaction, $F(6, 144) = 3.13, p = .01, \eta^2_p = .12$. The significant interaction required further analyses to examine the effect of time in each group and the effect of group at each testing point.

Results for the effect of time in each group

A repeated-measures ANOVA of the CR group’s scores for embedded questions in the UGJ tests found a significant difference over time, $F(2, 38) = 11.82, p = .00, \eta^2_p = .38$. Bonferroni post-hoc comparisons revealed the CR group significantly improved its scores from pretest to posttest 1, $p = .03, d = .63$, and from pretest to posttest 2, $p = .00, d = .98$. There was no significant difference between the posttests, $p > .05, d = .42$. 

Figure 11: Means for Embedded Questions in UGJ Tests over Time by Group
A repeated-measures ANOVA for the IR group also reported a significant main effect of time, $F(2, 48) = 13.12, p = .00, \eta^2_p = .35$. Bonferroni post-hoc analyses indicated the IR group’s scores in posttest 2 were significantly higher than in both the pretest, $p = .00, d = .93$, and posttest 1, $p = .00, d = .83$. There was no significant change between pretest and posttest 1, $p > .05, d = .03$.

A repeated-measures ANOVA for the task control group determined no significant main effect of time, $F(2, 28) = .40, p > .05, \eta^2_p = .03$, suggesting that the task control group’s accuracy in judging the rule of embedded questions did not change significantly from UGJ pretest to posttests.

A repeated-measures ANOVA for the test control group showed no significant difference over time either, $F(2, 30) = .85, p > .05, \eta^2_p = .05$.

**Results for the effect of group at each testing point**

A one-way ANOVA of the scores for embedded questions in the UGJ pretest found a significant effect of group, $F(3, 72) = 3.32, p = .02, \eta^2 = .12$. Bonferroni post-hoc analyses revealed the task control group had a marginally significantly higher pretest mean than the test control group, $p = .07$, with a large effect size, $d = .95$. Consequently, posttest scores were compared across groups using ANCOVAs controlling for the effect of pretest scores. Adjusted means for the UGJ posttests are presented in Table 54.

<table>
<thead>
<tr>
<th></th>
<th>Posttest 1</th>
<th>Posttest 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Std Error</td>
</tr>
<tr>
<td>CR group (n=20)</td>
<td>73.14</td>
<td>3.07</td>
</tr>
<tr>
<td>IR group (n=25)</td>
<td>66.33</td>
<td>2.75</td>
</tr>
<tr>
<td>TkC group (n=15)</td>
<td>63.67</td>
<td>3.58</td>
</tr>
<tr>
<td>TtC group (n=16)</td>
<td>65.89</td>
<td>3.48</td>
</tr>
</tbody>
</table>

Note. a. Covariates were evaluated at the value of 65.79.

CR = corrective recast; IR = implicit recast; TkC = task control; TtC = test control.

The ANCOVA for UGJ posttest 1 reported no significant difference among adjusted means for the four groups, $F(3, 71) = 1.63, p > .05, \eta^2_p = .06$. 158
The ANCOVA for UGJ posttest 2 showed a significant effect for group, $F(3, 71) = 4.21$, $p = .01$, $\eta_p^2 = .15$. Bonferroni post-hoc comparisons indicated that both the CR group ($p = .05, d = .76$) and the IR group ($p = .02, d = .78$) significantly outperformed the task control group in judging the rule of embedded questions. Other between-group differences in UGJ total scores were not significant.

**Table 55: A Summary of within- and between-Subject Differences in UGJ Tests for Embedded Questions**

<table>
<thead>
<tr>
<th></th>
<th>df</th>
<th>$p$ value</th>
<th>Effect size ($d$)</th>
<th>Effect size ($\eta^2$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Within-subject</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CR group: pretest - posttest 1</td>
<td>19</td>
<td>.03*</td>
<td>.63</td>
<td></td>
</tr>
<tr>
<td>CR group: pretest - posttest 2</td>
<td>19</td>
<td>.00***</td>
<td>.98#</td>
<td></td>
</tr>
<tr>
<td>CR group: posttest1 - posttest 2</td>
<td>19</td>
<td>.24</td>
<td>.42</td>
<td></td>
</tr>
<tr>
<td>IR group: pretest - posttest 1</td>
<td>24</td>
<td>1.00</td>
<td>.03</td>
<td></td>
</tr>
<tr>
<td>IR group: pretest - posttest 2</td>
<td>24</td>
<td>.00***</td>
<td>.93#</td>
<td></td>
</tr>
<tr>
<td>IR group: posttest1 - posttest 2</td>
<td>24</td>
<td>.00***</td>
<td>.83#</td>
<td></td>
</tr>
<tr>
<td>TkC group: omnibus</td>
<td>2</td>
<td>.68</td>
<td>.03</td>
<td></td>
</tr>
<tr>
<td>TkC group: omnibus</td>
<td>2</td>
<td>.44</td>
<td>.05</td>
<td></td>
</tr>
<tr>
<td>Pretest: CR - IR</td>
<td>43</td>
<td>.51</td>
<td>.52</td>
<td></td>
</tr>
<tr>
<td>Pretest: CR - TkC</td>
<td>33</td>
<td>.23</td>
<td>.78</td>
<td></td>
</tr>
<tr>
<td>Pretest: CR - TtC</td>
<td>34</td>
<td>1.00</td>
<td>.23</td>
<td></td>
</tr>
<tr>
<td>Pretest: IR - TkC</td>
<td>38</td>
<td>1.00</td>
<td>.20</td>
<td></td>
</tr>
<tr>
<td>Pretest: IR - TtC</td>
<td>39</td>
<td>.14</td>
<td>.70</td>
<td></td>
</tr>
<tr>
<td>Pretest: TkC - TtC</td>
<td>29</td>
<td>.07</td>
<td>.95#</td>
<td></td>
</tr>
<tr>
<td>Posttest 1: omnibus</td>
<td>3</td>
<td>.19</td>
<td></td>
<td>.06</td>
</tr>
<tr>
<td>Posttest 2: CR - IR</td>
<td>43</td>
<td>1.00</td>
<td>.01</td>
<td></td>
</tr>
<tr>
<td>Posttest 2: CR - TkC</td>
<td>33</td>
<td>.05*</td>
<td>.76</td>
<td></td>
</tr>
<tr>
<td>Posttest 2: CR - TtC</td>
<td>34</td>
<td>.35</td>
<td>.51</td>
<td></td>
</tr>
<tr>
<td>Posttest 2: IR - TkC</td>
<td>38</td>
<td>.02*</td>
<td>.78</td>
<td></td>
</tr>
<tr>
<td>Posttest 2: IR - TtC</td>
<td>39</td>
<td>.29</td>
<td>.53</td>
<td></td>
</tr>
<tr>
<td>Posttest 2: TkC - TtC</td>
<td>29</td>
<td>1.00</td>
<td>.24</td>
<td></td>
</tr>
</tbody>
</table>

Note. CR = corrective recast; IR = implicit recast; TkC = task control; TtC = test control.
* $p \leq .05$.  ** $p \leq .01$.  *** $p \leq .001$.  # = large effect size.

*A summary of significant differences in the UGJ total scores for embedded questions*

- The CR group’s total scores in the UGJ test for embedded questions significantly
improved from pretest to posttests.

- The IR group’s scores for embedded questions did not change significantly until posttest 2.

- The control groups’ scores in the UGJ test for embedded questions did not significantly change from pretest to posttests.

- The four groups had significantly different prior knowledge of embedded questions as indicated by the UGJ pretest scores.

- There was no significant between-group difference in UGJ posttest 1.

- Both the CF groups significantly outperformed the task control group in UGJ posttest 2.

### 6.2.6 Results for embedded questions: Untimed grammaticality judgment (ungrammatical items)

**Descriptive statistics**

Table 56 presents each group’s ungrammatical item means and SDs in the untimed grammaticality judgment tests for embedded questions. The CR group had a mean score of 47.78 in the pretest and obtained a higher mean of 55.19 in posttest 1 and an even higher mean of 59.81 in posttest 2 (see Figure 12). The IR group was stable from pretest (M = 52.68) to posttest 1 (M = 53.64) but had a sudden growth from posttest 1 to posttest 2 (M = 71.07). The task control group’s means fluctuated from 55.23 in the pretest down to 51.96 in posttest 1, and subsequently up to 54.90 in posttest 2. The test control group’s scores crept up from 39.22 in the pretest to 43.14 in posttest 1 and then to 47.71 in posttest 2.
Table 56: Descriptive Statistics of UGJ (Ungrammatical Item) Scores for Embedded Questions

<table>
<thead>
<tr>
<th></th>
<th>Pretest</th>
<th></th>
<th>Posttest 1</th>
<th></th>
<th>Posttest 2</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>CR group (n=30)</td>
<td>47.78</td>
<td>24.61</td>
<td>55.19</td>
<td>27.75</td>
<td>59.81</td>
<td>27.51</td>
</tr>
<tr>
<td>IR group (n=29)</td>
<td>52.68</td>
<td>21.91</td>
<td>56.40</td>
<td>28.40</td>
<td>71.07</td>
<td>26.66</td>
</tr>
<tr>
<td>TkC group (n=17)</td>
<td>55.23</td>
<td>19.21</td>
<td>51.96</td>
<td>25.30</td>
<td>54.90</td>
<td>30.47</td>
</tr>
<tr>
<td>TtC group (n=17)</td>
<td>39.22</td>
<td>24.41</td>
<td>43.14</td>
<td>22.70</td>
<td>47.71</td>
<td>30.74</td>
</tr>
<tr>
<td>Total (n=93)</td>
<td>49.10</td>
<td>23.12</td>
<td>51.91</td>
<td>26.61</td>
<td>60.21</td>
<td>29.19</td>
</tr>
</tbody>
</table>

Note. CR = corrective recast; IR = implicit recast; TkC = task control; TtC = test control.

Figure 12: Means for Embedded Questions in UGJ (Ungrammatical Item) Tests over Time by Group

Inferential statistics

Omnibus results

A mixed ANOVA of the ungrammatical item scores for embedded questions in the UGJ tests
determined there was a marginally significant effect for the time x group interaction, \( F(6, 178) = 2.00, p = .07, \eta^2_p = .06 \). This indicated that further analyses should be conducted to examine the effect of time in each group and the effect of group at each testing point.

**Results for the effect of time in each group**

A repeated-measures ANOVA of the CR group’s ungrammatical item scores for embedded questions in the UGJ tests found a significant difference over time, \( F(2, 58) = 4.46, p = .02, \eta^2_p = .13 \). Bonferroni post-hoc comparisons revealed the CR group did not significantly improve its scores for ungrammatical items until posttest 2, \( p = .02, d = .54 \).

A repeated-measures ANOVA for the IR group also reported a significant main effect of time, \( F(2, 56) = 12.69, p = .00, \eta^2_p = .31 \). Bonferroni post-hoc analyses indicated the IR group’s scores in posttest 2 were significantly higher than in the pretest, \( p = .00, d = .86 \), and in posttest 1, \( p = .00, d = .70 \). There was no significant change from pretest to posttest 1, \( p > .05, d = .05 \).

A repeated-measures ANOVA for the task control group determined no significant main effect of time, \( F(2, 32) = .19, p > .05, \eta^2_p = .01 \), suggesting that the task control group’s accuracy in judging the rule of embedded questions did not change significantly from UGJ pretest to posttests.

A repeated-measures ANOVA for the test control group’s ungrammatical item scores showed no significant difference over time either, \( F(2, 32) = 1.50, p > .05, \eta^2_p = .09 \).

**Results for the effect of group at each testing point**

A one-way ANOVA of the UGJ ungrammatical item pretest scores showed that the four groups’ means for embedded questions did not differ significantly prior to the treatment, \( F(3, 89) = 1.74, p > .05, \eta^2 = .06 \).

A one-way ANOVA of the scores for ungrammatical items in UGJ posttest 1 found no significant difference among the four groups, \( F(3, 89) = .80, p > .05, \eta^2 = .03 \).

A one-way ANOVA for UGJ posttest 2 reported that there were between-group differences, \( F(3, 89) = 2.71, p = .05, \eta^2 = .08 \). Bonferroni post-hoc comparisons indicated that the IR group significantly outperformed the test control group in judging the rule of embedded questions, \( p \)
= .05, \(d = .81\). Other between-group differences in UGJ posttest 2 were not significant.

### Table 57: A Summary of within- and between-Subject Differences in UGJ (Ungrammatical Item) Tests for Embedded Questions

<table>
<thead>
<tr>
<th></th>
<th>df</th>
<th>(p) value</th>
<th>Effect size ((d))</th>
<th>Effect size ((\eta^2))</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Within-subject</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CR group: pretest - posttest 1</td>
<td>29</td>
<td>.35</td>
<td>.30</td>
<td></td>
</tr>
<tr>
<td>CR group: posttest - posttest 2</td>
<td>29</td>
<td>.02*</td>
<td>.54</td>
<td></td>
</tr>
<tr>
<td>CR group: posttest 1 - posttest 2</td>
<td>29</td>
<td>.57</td>
<td>.25</td>
<td></td>
</tr>
<tr>
<td><strong>Between-subject</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pretest: omnibus</td>
<td>3</td>
<td>.17</td>
<td>.06</td>
<td></td>
</tr>
<tr>
<td>Posttest 1: omnibus</td>
<td>3</td>
<td>.50</td>
<td>.03</td>
<td></td>
</tr>
<tr>
<td>Posttest 2: CR - IR</td>
<td>57</td>
<td>.79</td>
<td>.42</td>
<td></td>
</tr>
<tr>
<td>Posttest 2: CR - TkC</td>
<td>45</td>
<td>1.00</td>
<td>.17</td>
<td></td>
</tr>
<tr>
<td>Posttest 2: CR - TtC</td>
<td>45</td>
<td>.99</td>
<td>.41</td>
<td></td>
</tr>
<tr>
<td>Posttest 2: IR - TkC</td>
<td>44</td>
<td>.39</td>
<td>.56</td>
<td></td>
</tr>
<tr>
<td>Posttest 2: IR - TtC</td>
<td>44</td>
<td>.05*</td>
<td>.81#</td>
<td></td>
</tr>
<tr>
<td>Posttest 2: TkC - TtC</td>
<td>32</td>
<td>1.00</td>
<td>.23</td>
<td></td>
</tr>
</tbody>
</table>

Note. CR = corrective recast; IR = implicit recast; TkC = task control; TtC = test control.

* \(p \leq .05\). ** \(p \leq .01\). *** \(p \leq .001\). \# = large effect size.

A summary of significant differences in the UGJ ungrammatical item scores for embedded questions

- The CR group’s ungrammatical item scores for embedded questions in UGJ posttest 2 were significantly higher than in UGJ pretest.
- The IR group’s improvement from UGJ posttest 1 to posttest 2 was significant.
- The two control groups’ ungrammatical item scores did not change significantly from UGJ pretest to UGJ posttests.
- There were no significant between-group differences in UGJ posttest 1 considering...
only ungrammatical items.

- The IR group outperformed the test control group in UGJ posttest 2 when only ungrammatical items for embedded questions were analyzed.

6.3 Discussion

Research Question 2 concerns the link between recast types and the learning of explicit knowledge as measured by the written production and untimed grammaticality judgment (ungrammatical item) tests, examining whether explicitness contributes to the benefit of recasts in terms of explicit knowledge.

Question 2a asks whether corrective recasts facilitate L2 explicit knowledge development. The answer is affirmative. The CR group was significantly more accurate in the written production and grammaticality judgment of 3rd person -s and embedded questions in posttests than in the pretest. The CR group not only outperformed the test control group in UGJ (ungrammatical item) posttests for 3rd person -s, but also outperformed the task control group significantly in UGJ posttest 1 for 3rd person -s (involving intermediate and high proficiency learners) and UGJ posttest 2 for embedded questions. As discussed in Chapter 5, the benefits of CR could be due to the fact that corrective recasts consisted of two feedback moves: a repetition with rising intonation and a recast with falling intonation. The prosodic and grammatical juxtaposition of learners’ erroneous utterance and the teacher’s targetlike correction may have enhanced the salience of target structures. The salience in turn made it possible for the learner to notice the gap, which triggered subsequent L2 learning. In line with Doughty and Varela (1998) and Doughty (2001), this research supports the claim that attention-getters preceding recasts eliminate the potential ambiguity of recasting, thus rendering the CF salient and facilitative of L2 development (Loewen & Philp, 2006; Nassaji, 2009; Sheen, 2006).

The finding that recasts are effective in the UGJ test is in stark contrast to Révész’s (2012) study, which reported the lowest gains in the UGJ test. She argued that the oral CF treatment leads to more development of implicit knowledge as measured by oral production tests than explicit knowledge measured by written tests according to the Transfer-Appropriate Processing Hypothesis (TAP). The knowledge acquired in an oral communicative context is
difficult to retrieve in the written production test and more difficult to deploy in the
grammaticality judgment test, which depends on different operating conditions. However, the
current study’s findings regarding the effect of corrective recasts in the UGJ tests suggest that
explicit recasts significantly promote the development of explicit knowledge in addition to
implicit knowledge (as shown in Chapter 5). It can be speculated that the CF has facilitated
the transfer from oral skills to written skills through explicit knowledge.

The inefficacy of CF and oral interaction in the WP tests could stem from a ceiling effect. A
total mean score of 74.69 in the pretest for 3rd person -s and 76.02 for embedded questions
indicates that the participants had already obtained sufficient explicit knowledge of the target
structures prior to the treatment, and there would be little space for them to develop. As a
result, all the groups performed equally well in posttests as illustrated by the CR, IR and task
control groups’ significant improvement from WP pretest to Posttests but no significant
between-group differences were observed in this study. The WP test that this study used for
both target structures did not constitute an efficient measure of the effect of recasts on the
development of explicit knowledge. A proper WP test may report a positive effect for explicit
recasts, as shown in Mifka-Profozic’s (2013) study. Her recast group outperformed the
control and clarification request groups in the written production delayed posttest for passé
composé and imparfait.

Question 2b addresses whether implicit recasts are effective in promoting L2 explicit
knowledge development. The answer is also affirmative. The IR group significantly improved
its scores for 3rd person -s in the WP and UGJ immediate posttests and for embedded
questions in the UGJ delayed posttest. The IR group outperformed not only the test control
group in the UGJ (ungrammatical item) posttests but also the task control group in UGJ
posttest 1 for 3rd person -s (involving high proficiency learners) and in UGJ posttest 2 for
embedded questions. The effectiveness of implicit recasts on L2 explicit knowledge could be
attributed to feedback contingency, feedback intensity and learner orientation. As Kim (2004)
and Long (2007) argued, recasts are contingent on learners’ errors and are given immediately
following learners’ output. The learners can free their attentional resources from meaning to
form. They may have utilized the time-out opportunity to develop explicit knowledge. The
second contributing factor is feedback intensity. The teacher gave 79 implicit recasts
consistently on 3rd person -s errors during the three sessions in this research. The implicit
recasts were so intensive that they could have served as positive evidence even if the
corrective intent was not identified. It was likely that the learners drew on the exemplars of the target structure in the CF to enhance their explicit knowledge. The explicit knowledge was improved immediately in the case of 3rd person -s but in the long term in the case of embedded questions due to the latter structure’s complexity.

The finding that implicit recasts were beneficial for L2 learning when the intensity condition is met is consistent with Han’s (2002) laboratory-based study, but not with Lyster’s (2004) classroom-based study. The discrepancy could possibly result from the difference in contexts. Lyster’s study was conducted in a Canadian French immersion setting, whereas this research occurred in an EFL setting, where the teacher and learners had a form-focused orientation towards classroom interaction. The learners were struggling to produce grammatically accurate utterances by repeating what they said. Implicit recasts in a Chinese EFL setting, which constitute a counterbalance to learners’ emphasis on form, are likely to be effective for L2 learning (Lyster & Mori, 2006). That is to say, learners’ form-orientation assisted them to recognize the teacher’s corrective intent when implicit recasts were provided. The successful recognition of the linguistic form that the implicit recasts targeted may have enhanced the benefit of the CF.

This finding supports Goo and Mackey’s claim (2013) that recasts work for L2 learning in classroom settings. Lyster and Ranta (2013) questioned this claim by pointing out a methodological deficiency in some classroom studies that investigated the recast-learning relationship. They suggested that the control group should be exposed to the same amount of activities as treatment groups so that the effect of CF can be examined separately. Using such a comparison group, the current research found that the IR group outperformed the task control group, indicating that implicit recasts play a facilitative role in L2 grammar learning.

Question 2c examines whether there is any difference in the effect of corrective recasts and implicit recasts on L2 explicit knowledge. The answer appears to be affirmative. The CR group outperformed the IR group in UGJ ungrammatical item posttest 2 for 3rd person -s when low proficiency learners were involved. The finding that corrective recasts outweighed implicit recasts in this research contradicts Erlam and Loewen’s (2010) study. In their laboratory setting, four members per group on average received implicit recasts more intensively. Their implicit recasts may be as salient as explicit recasts, leading to no differentiated effects between the CF types. Although the study reported in this thesis occurred in an EFL classroom, where the students’ form-orientation and the intensity of the
CF rendered the target structure more salient, there was still a considerable difference in salience between the explicit and implicit recasts, as demonstrated by the significant difference in uptake-with-repair rate (see Chapter 7).

It should be noted that the CR and IR groups did not differ significantly in most of the analyses. Their potential difference might have been overwhelmed by the significant effect of oral interaction. As the results reveal, the task control group significantly outperformed the test control group in UGJ posttest 1 for 3rd person -s. One possible explanation for the significant effect of interaction is task features. All the treatment tasks were designed to elicit a certain grammatical structure, that is to say, learners could not have completed the tasks without using the grammatical structure. The repeated use of the target structure in the tasks might have oriented the task control group’s attention to the form of 3rd person -s (or embedded questions) even though it did not receive any CF. The effectiveness of oral interaction alone in this study is in line with Erlam and Loewen’s (2010) finding and Mackey and Goo’s (2007) meta-analysis results.

It is an interesting finding that the effect of recasts varied with learner proficiency. For those high proficiency learners, both corrective recasts and implicit recasts were effective as illustrated by the CR and IR groups’ higher scores for 3rd person -s in UGJ posttest 1 than the task control group. For those medium proficiency learners, only corrective recasts were effective as shown by the CR group’s superiority over the task control group. For those low proficiency learners, neither the CR group nor the IR group outperformed the task control group in their ability to judge the rule of 3rd person -s. This finding, consistent with those in Ammar (2008) and Mifka-Profozic (2013), lends support to the claim that the effectiveness of recasts is subject to learners’ proficiency level (Ammar & Spada, 2006). High proficiency learners appear to benefit equally from explicit and implicit recasts because they can afford to spare attentional resources to the feedback while keeping their main focus on meaning. In contrast, low proficiency learners are likely to be constrained from noticing the feedback by their proficiency level. It was evident that low proficiency learners were less accurate in recalling their noticing of recasts in Philp (2003). When learners have difficulty in identifying the target structure of recasts, explicitness in the CF makes a difference. Explicit recasts are more likely to draw the low-proficiency learners’ attention to linguistic form leading to more L2 learning than implicit recasts. This explains why corrective recasts were more effective than implicit recasts only for low proficiency learners in UGJ ungrammatical item posttest 2.
for 3rd person -s.

6.4 Summary
In summary, corrective recasts facilitated the learners’ explicit knowledge as measured by UGJ tests due to the immediate juxtaposition of error repetition and correct recasting with contrasting intonation, which led to noticing the gap. Implicit recasts were also beneficial for L2 explicit knowledge development because of their contingency on learner errors, their intensity while targeting one single structure, and the learners’ form-focused orientation in a foreign language context. These conditions made attention to form possible although the CF was implicit. Corrective recasts’ superiority over implicit recasts was dependent on learners’ proficiency level. High proficiency learners can afford attention to form, notice the gap, and finally benefit from recasts, irrespective of CF explicitness. Low proficiency learners benefited little from recasts due to the constraint of the availability of attentional resources, when explicit recasts had an advantage over implicit recasts in triggering noticing of the CF and thus leading to L2 learning. Oral interaction alone facilitated the development of explicit knowledge, which overwhelmed the effect of recasts, the implicit type in particular.
Chapter 7 Uptake and Acquisition: Results and Discussion

Chapter 7 addresses two research questions: one concerns whether there is a difference in uptake-with-repair following corrective and implicit recasts and the other examines the link between repair rate and subsequent L2 learning.

**RQ3:** Is there any difference in uptake-with-repair following corrective recasts and implicit recasts?

**RQ4:** Is there any relationship between learners’ uptake-with-repair and their subsequent acquisition of implicit and explicit knowledge?

### 7.1 Statistical treatment

Learners’ repetition or incorporation of a recast was coded as “repair”. As a subcategory of repair, “repetition” refers to a parrot-like utterance that simply repeats the correct form provided by the recast, and “incorporation” is a learner response that integrates the correct form into a longer utterance. Examples of these two types of repair can be found in Note 2 at the end of the thesis. Following the recast, a learner might also produce no response, continue the topic without taking the opportunity to focus on form, acknowledge the teacher’s feedback, or attempt to repair but fail. All these learner moves, which produced no correct form, were coded as “non-repair”.

To answer Question 3, the corrective recast (CR) and implicit recast (IR) groups’ repair and non-repair turns were tallied and a Pearson chi-square ($\chi^2$) test of the frequencies was conducted. Similarly, the frequencies of those participants who successfully repaired (repairers) and who failed to repair their errors (non-repairers) in the CR and IR groups were compared via another chi-square test. The effect size (ES) of chi-square tests is reported using a phi ($\phi$) value, in which .1 to .3 is small ES, .3 to .5 is medium ES, and .5 to 1 is large ES.

To answer Question 4, each feedback receiver’s repair rate was calculated using a formula: repair frequency divided by the total number of received recasts. Language development took the form of gain scores for implicit/explicit knowledge from pretest to posttest. As justified in
Chapter 5 and Chapter 6, elicited imitation (EI) and oral production (OP) scores were synthesized into implicit knowledge scores, and written production (WP) and untimed grammaticality judgment (UGJ) ungrammatical item scores were combined into explicit knowledge scores. For the CR group, bivariate correlations were conducted to see whether there was a relationship between repair rate and gain scores for implicit/explicit knowledge. For the IR group, however, linear correlation analyses were not appropriate because more than two thirds of participants did not repair their errors and consequently the data had a nonlinear distribution. Therefore, the IR group’s repairers and non-repairers’ gain scores were compared using independent-samples \( t \) tests. The effect size of \( t \) tests is reported using a Cohen’s \( d \) value, in which .2 to .5 is small ES, .5 to .8 is medium ES, and any value above .8 is large ES. Finally, these analyses were performed again using incorporation rates instead of repair rates to see whether incorporation was a better predictor of L2 learning.

It should be noted that the sample size varies from analysis to analysis. Those participants who did not receive feedback and those feedback receivers who scored higher than 90 in the pretest were excluded from the analyses. Only those participants who had repair and language scores available were included. Outliers were also deleted.

### 7.2 Results for uptake-with-repair

#### 7.2.1 Results for repair frequency: 3rd person -s

Study 1 targeted 3rd person -s. For this target structure, altogether 80 corrective recasts were provided, 51 of which led to uptake-with-repair in the CR group. The IR group received 79 implicit recasts but only 11 of them were followed by repair moves (see Table 58).

<table>
<thead>
<tr>
<th></th>
<th>Uptake-with-repair</th>
<th>Non-repair</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>CR group (( n = 24^a ))</td>
<td>51</td>
<td>29</td>
<td>80</td>
</tr>
<tr>
<td>IR group (( n = 26^a ))</td>
<td>11</td>
<td>68</td>
<td>79</td>
</tr>
</tbody>
</table>

*Note. a. Sample size of feedback receivers. CR = corrective recast; IR = implicit recast.*

A chi-square test showed that there was a statistically significant association between the type of corrective feedback (CF) and the frequency of learner repair, \( \chi^2 \) (1, 159) = 41.48, \( p = .00, \varphi \)
= .51 (large ES). In other words, the extent to which the participants repaired their 3rd person -s errors depended on the type of CF they received. Corrective recasts were more likely to result in uptake-with-repair than implicit recasts, as illustrated by the odds ratio of 10.9.

7.2.2 Results for repair frequency: Embedded questions

Study 2 targeted embedded questions. In Study 2, the CR group participated in 90 CF episodes (see Table 59), 33 of which ended with uptake-with-repair. The IR group received 129 implicit recasts but only two of them were followed by repair moves.

<table>
<thead>
<tr>
<th></th>
<th>Uptake-with-repair</th>
<th>Non-repair</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>CR group (n = 24&lt;sup&gt;a&lt;/sup&gt;)</td>
<td>33</td>
<td>57</td>
<td>90</td>
</tr>
<tr>
<td>IR group (n = 28&lt;sup&gt;a&lt;/sup&gt;)</td>
<td>2</td>
<td>127</td>
<td>129</td>
</tr>
</tbody>
</table>

Note. <sup>a</sup> Sample size of feedback receivers.

CR = corrective recast; IR = implicit recast.

A chi-square test indicated a significant difference in uptake-with-repair between the CR and IR groups, $\chi^2 (1, 219) = 48.69, p = .00, \phi = .47$. The odds ratio confirmed that the odds of learner repair were 36.8 times higher following corrective recasts than those following implicit recasts for embedded questions.

7.2.3 Results for repairers vs. non-repairers: 3rd person -s

In Study 1, altogether 24 participants received corrective recasts on 3rd person -s, and 21 of them repaired their errors following the recasts. 26 participants were corrected by means of implicit recasts, but only seven of them repaired their errors (see Table 60).

<table>
<thead>
<tr>
<th></th>
<th>Repairers</th>
<th>Non-Repairers</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>CR group</td>
<td>21</td>
<td>3</td>
<td>24</td>
</tr>
<tr>
<td>IR group</td>
<td>7</td>
<td>19</td>
<td>26</td>
</tr>
</tbody>
</table>

Note. CR = corrective recast; IR = implicit recast.
A chi-square test indicated a significant difference between the number of learners who repaired 3\textsuperscript{rd} person -s errors following corrective recasts and the number who repaired following implicit recasts, $\chi^2 (1, 50) = 18.59, p = .00, \phi = .61$ (large ES). Corrective recasts were more likely to lead to repair than implicit recasts as illustrated by the odds ratio of 19.0.

### 7.2.4 Results for repairers vs. non-repairers: Embedded questions

In Study 2, altogether 24 participants received corrective recasts on embedded questions, and 19 of them produced uptake-with-repair. Among the 28 participants who were corrected by implicit recasts, only two of them repaired their errors (see Table 61).

| Table 61: Repairers and Non-Repairers for Embedded Questions |
|---------------|---------|---------|
| CR group      | 19      | 5       | 24      |
| IR group      | 2       | 26      | 28      |

Note. CR = corrective recast; IR = implicit recast.

A chi-square test for Study 2 determined that the difference in the number of repairers in the two feedback conditions was statistically significant, $\chi^2 (1, 52) = 27.85, p = .00, \phi = .73$ (large ES). As the odds ratio showed, corrective recasts were 49.4 times more likely to result in repair than implicit recasts on embedded questions.

### 7.3 Results for the repair-acquisition relationship

The relationship between repair rate and knowledge gains was examined by means of bivariate correlations in the CR group, but independent-samples $t$ tests of the repairer and non-repairer subgroups’ gain scores in the IR group due to the data’s nonlinear distribution. The results are reported for the two target structures separately.

#### 7.3.1 The CR group’s repair-acquisition relationship for 3\textsuperscript{rd} person -s

Study 1 targeted 3\textsuperscript{rd} person -s. After excluding those learners who scored over 90 in the EI and OP pretests and one feedback receiver who missed posttests, the CR group had 23 valid participants for the correlation analyses between repair and gains in implicit knowledge. After deleting those scoring higher than 90 in the WP and UGJ pretests, there were only 19 valid participants for the repair - explicit knowledge correlations. The CR group’s sample size, as
well as the means for repair rate and gain scores in each analysis is presented in Table 62.

Table 62: The CR Group’s Repair and Gain Scores for 3rd Person -s

<table>
<thead>
<tr>
<th></th>
<th>Repair</th>
<th>Gain 1</th>
<th>Gain 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
</tr>
<tr>
<td>Repair – Implicit (n = 23)</td>
<td>.65</td>
<td>.33</td>
<td>19.04</td>
</tr>
<tr>
<td>Repair – Explicit (n = 19)</td>
<td>.59</td>
<td>.33</td>
<td>18.24</td>
</tr>
</tbody>
</table>

Note. Gain 1 = gain score from pretest to immediate posttest; Gain 2 = gain score from pretest to delayed posttest.
Implicit = implicit knowledge; explicit = explicit knowledge.

Bivariate correlations determined that repair rate following corrective recasts on 3rd person -s was not significantly correlated with gain scores in either implicit knowledge or explicit knowledge (see Table 63).

Table 63: Correlations for the CR Group’s Repair and Gain Scores for 3rd Person -s

<table>
<thead>
<tr>
<th></th>
<th>Uptake-with-repair</th>
</tr>
</thead>
<tbody>
<tr>
<td>Implicit knowledge gain 1 (n = 23)</td>
<td>.20</td>
</tr>
<tr>
<td>Implicit knowledge gain 2 (n = 23)</td>
<td>-.19</td>
</tr>
<tr>
<td>Explicit knowledge gain 1 (n = 19)</td>
<td>.18</td>
</tr>
<tr>
<td>Explicit knowledge gain 2 (n = 19)</td>
<td>-.06</td>
</tr>
</tbody>
</table>

Note. *p < .05. **p < .01.

7.3.2 The CR group’s repair-acquisition relationship for embedded questions
In Study 2, which targeted embedded questions, the CR group had 24 valid participants for the correlations between repair and implicit knowledge gains, and 23 participants for the correlations between repair and explicit knowledge gains after excluding those feedback receivers who scored over 90 in the pretests. The CR group's sample size, as well as the means for repair rate and knowledge gain scores is reported for each analysis in Table 64.
Table 64: The CR group’s Repair and Gain Scores for Embedded Questions

<table>
<thead>
<tr>
<th>Repair – Implicit (n = 24)</th>
<th>Gain 1 Mean</th>
<th>Gain 1 SD</th>
<th>Gain 2 Mean</th>
<th>Gain 2 SD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>.45</td>
<td>.39</td>
<td>10.58</td>
<td>14.72</td>
</tr>
<tr>
<td>Repair – Explicit (n = 23)</td>
<td>.44</td>
<td>.40</td>
<td>3.89</td>
<td>24.18</td>
</tr>
</tbody>
</table>

Note. Gain 1 = gain score from pretest to immediate posttest; Gain 2 = gain score from pretest to delayed posttest. Implicit = implicit knowledge; explicit = explicit knowledge.

Bivariate correlations reported that the repair rate following corrective recasts on embedded questions was significantly positively correlated with the development of implicit knowledge, \( r (24) = .43, p = .04 \), but was significantly negatively associated with the development of explicit knowledge, \( r (23) = -.44, p = .04 \) in the immediate posttest (see Table 65). Although the directions were consistent, the significantly positive and negative correlations were only with the immediate gains, but not with the long-term gains.

Table 65: Correlations for the CR Group’s Repair and Gain Scores for Embedded Questions

<table>
<thead>
<tr>
<th></th>
<th>Uptake-with-repair</th>
</tr>
</thead>
<tbody>
<tr>
<td>Implicit knowledge gain 1 (n = 24)</td>
<td>.43*</td>
</tr>
<tr>
<td>Implicit knowledge gain 2 (n = 24)</td>
<td>.21</td>
</tr>
<tr>
<td>Explicit knowledge gain 1 (n = 23)</td>
<td>-.44*</td>
</tr>
<tr>
<td>Explicit knowledge gain 2 (n = 23)</td>
<td>-.20</td>
</tr>
</tbody>
</table>

Note. *p < .05. **p < .01.

7.3.3 The IR group’s repair-acquisition relationship for 3rd person -s

Following implicit recasts on 3rd person -s, which Study 1 targeted, only seven participants repaired their errors. Therefore, the IR group’s repair-acquisition relationship was investigated by comparing repairers and non-repairers’ gain scores. The IR repair and non-repair subgroups’ sample sizes and gain scores in implicit/explicit knowledge are presented in Table 66. The repairer subgroup’s mean gains are lower than the non-repairers’. Note that the repairers’ mean for explicit knowledge gains was close to 0 in the immediate posttest and negative in the delayed posttest.
Table 66: The IR Repair and Non-Repair Subgroups’ Gain Scores for 3rd Person -s

<table>
<thead>
<tr>
<th></th>
<th>Implicit gain 1</th>
<th>Implicit gain 2</th>
<th>Explicit gain 1</th>
<th>Explicit gain 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>Repair subgroup</td>
<td>9.00</td>
<td>11.73</td>
<td>8.44</td>
<td>10.63</td>
</tr>
<tr>
<td>(n = 7 / 4a)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-Repair subgroup</td>
<td>9.56</td>
<td>13.19</td>
<td>9.28</td>
<td>12.25</td>
</tr>
<tr>
<td>(n = 19 / 13b)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. IR = implicit recasts; Implicit = implicit knowledge; Explicit = explicit knowledge.
Gain 1 = gain score from pretest to immediate posttest; gain 2 = gain score from pretest to delayed posttest.
a. The repair subgroup’s sample size was 7 for implicit knowledge gains and 4 for explicit knowledge gains.
b. The non-repair subgroup’s sample size was 19 for implicit knowledge gains and 13 for explicit knowledge gains.

Independent-samples \( t \) tests revealed significant between-subgroup differences only in explicit knowledge gains. The IR non-repairers outperformed the IR repairers in the immediate posttest, \( t (15) = 2.16, p = .05 \), and in the delayed posttest, \( t (15) = 2.38, p = .03 \) (see Table 67). The results indicated that repair following implicit recasts on 3rd person -s errors was negatively related to the development of explicit knowledge.

Table 67: T-tests of the IR Repair and Non-Repair Subgroups’ Gain Scores for 3rd Person -s

<table>
<thead>
<tr>
<th></th>
<th>df</th>
<th>( t )</th>
<th>Sig</th>
</tr>
</thead>
<tbody>
<tr>
<td>Implicit knowledge gain 1</td>
<td>24</td>
<td>.10</td>
<td>.92</td>
</tr>
<tr>
<td>Implicit knowledge gain 2</td>
<td>24</td>
<td>.16</td>
<td>.87</td>
</tr>
<tr>
<td>Explicit knowledge gain 1</td>
<td>15</td>
<td>2.16</td>
<td>.05*</td>
</tr>
<tr>
<td>Explicit knowledge gain 2</td>
<td>15</td>
<td>2.38</td>
<td>.03*</td>
</tr>
</tbody>
</table>

Note. IR = implicit recasts. *p < .05. **p < .01.

7.3.4 The IR group’s repair-acquisition relationship for embedded questions

In Study 2, which targeted embedded questions, only two out of 28 feedback receivers produced uptake-with-repair following implicit recasts. The repair subgroup’s extremely small sample size did not allow any inferential analysis.
7.3.5 A summary of repair-acquisition correlations

- Repair following CR on embedded question errors was positively correlated with the implicit knowledge gains in the immediate posttest.
- Repair following CR on embedded question errors was negatively associated with the explicit knowledge gains in the immediate posttest.
- Repair following IR on 3rd person -s errors was negatively related to the explicit knowledge gains in the immediate posttest.
- Repair following IR on 3rd person -s errors was negatively related to the explicit knowledge gains in the delayed posttest.

Taken together, repair was positively correlated with the development of implicit knowledge but negatively with explicit knowledge in the corrective recast condition; it was only negatively associated with the development of explicit knowledge in the implicit recast condition. The correlations between repair rate and knowledge gains were found to be negative by and large, especially in terms of explicit knowledge.

7.4 Results for the incorporation-acquisition relationship

Learners repair their errors following corrective or implicit recasts in two ways, either they simply repeat the correct form provided by the feedback or they incorporate the correct form into a longer utterance (see Note 2). A simple repetition may be parrot-like because it does not necessarily require any cognitive processing on the part of learners (Gass, 2003) while incorporation arguably involves greater processing as it integrates the correct form. Therefore, incorporation as a subcategory of repair was tallied for each feedback receiver and the incorporation rate was calculated based on the total recasts received. The incorporation-acquisition relationship was investigated by means of bivariate correlations for the CR group, but independent-samples t tests of repairers and non-repairers’ gain scores for the IR group.

7.4.1 The CR group’s incorporation-acquisition relationship for 3rd person -s

In Study 1, which targeted 3rd person -s, the CR group had 23 valid participants for the correlation analyses between incorporation rate and implicit knowledge gains and 19 participants for the correlation with explicit knowledge gains as a result of deleting the feedback receiver missing posttests and those scoring over 90 in the pretest. The CR group’s sample size for each analysis, as well as the means for incorporation rate and gain scores, is
shown in Table 68.

### Table 68: The CR Group’s Incorporation and Gain Scores for 3rd Person -s

<table>
<thead>
<tr>
<th>Incorporation</th>
<th>Gain 1</th>
<th>Gain 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
</tr>
<tr>
<td>Incorporation – Implicit (n = 23)</td>
<td>.50</td>
<td>.35</td>
</tr>
<tr>
<td>Incorporation – Explicit (n = 19)</td>
<td>.43</td>
<td>.35</td>
</tr>
</tbody>
</table>

Note. Gain 1 = gain score from pretest to immediate posttest; Gain 2 = gain score from pretest to delayed posttest. Implicit = implicit knowledge; explicit = explicit knowledge.

Bivariate correlation analyses indicated that incorporation following corrective recasts on 3rd person -s errors was not significantly related to any gain scores (see Table 69).

### Table 69: Correlations for the CR Group’s Incorporation and Gain Scores for 3rd Person -s

<table>
<thead>
<tr>
<th>Incorporation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Implicit knowledge gain 1 (n = 23)</td>
</tr>
<tr>
<td>Implicit knowledge gain 2 (n = 23)</td>
</tr>
<tr>
<td>Explicit knowledge gain 1 (n = 19)</td>
</tr>
<tr>
<td>Explicit knowledge gain 2 (n = 19)</td>
</tr>
</tbody>
</table>

Note. *p < .05. **p < .01.

### 7.4.2 The CR group’s incorporation-acquisition relationship for embedded questions

In Study 2, which targeted embedded questions, the CR group’s sample size was 24 for the correlations between incorporation rate and implicit knowledge gains, and 23 for the correlations with explicit knowledge gains after excluding those feedback receivers who scored over 90 in the pretest. The CR group’s sample size for each analysis, together with the means for incorporation rate and gains scores, is presented in Table 70.
Table 70: The CR Group’s Incorporation and Gain Scores for Embedded Questions

<table>
<thead>
<tr>
<th>Incorporation</th>
<th>Gain 1</th>
<th>Gain 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>Incorporation – Implicit (n = 24)</td>
<td>.28</td>
<td>.34</td>
</tr>
<tr>
<td>Incorporation – Explicit (n = 23)</td>
<td>.27</td>
<td>.35</td>
</tr>
</tbody>
</table>

Note. Gain 1 = gain score from pretest to immediate posttest; Gain 2 = gain score from pretest to delayed posttest. Implicit = implicit knowledge; explicit = explicit knowledge.

As the bivariate correlations indicated, incorporation rate was found to be significantly correlated with implicit knowledge gains for embedded questions in the immediate posttest, $r(24) = .55$, $p = .01$, and in the delayed posttest, $r(24) = .48$, $p = .02$, but not with explicit knowledge gains (see Table 71).

Table 71: Correlations for the CR Group’s Incorporation and Gain Scores for Embedded Questions

<table>
<thead>
<tr>
<th>Incorporation</th>
<th>Gain 1</th>
<th>Gain 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Implicit knowledge gain 1 (n = 24)</td>
<td>.55**</td>
<td></td>
</tr>
<tr>
<td>Implicit knowledge gain 2 (n = 24)</td>
<td>.48*</td>
<td></td>
</tr>
<tr>
<td>Explicit knowledge gain 1 (n = 23)</td>
<td>-.18</td>
<td></td>
</tr>
<tr>
<td>Explicit knowledge gain 2 (n = 23)</td>
<td>-.05</td>
<td></td>
</tr>
</tbody>
</table>

Note. *p < .05. **p < .01.

7.4.3 The IR group’s incorporation-acquisition relationship for 3rd person -s

Following implicit recasts on 3rd person -s, only six participants incorporated the correct form into their utterances at least once. Therefore, the IR group’s incorporation-acquisition relationship was investigated by comparing incorporators and non-incorporators’ gain scores by means of independent-samples $t$ tests. The IR incorporation and non-incorporation subgroups’ sample sizes and the means for incorporation rate and gain scores are presented in Table 72.
Table 72: The IR Incorporation and Non-Incorporation Subgroups’ Gain Scores for 3rd Person -s

<table>
<thead>
<tr>
<th></th>
<th>Implicit gain 1</th>
<th></th>
<th>Implicit gain 2</th>
<th></th>
<th>Explicit gain 1</th>
<th></th>
<th>Explicit gain 2</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>Incorporation subgroup (n = 6/3a)</td>
<td>7.64</td>
<td>12.23</td>
<td>9.99</td>
<td>10.74</td>
<td>3.28</td>
<td>16.61</td>
<td>-4.35</td>
<td>16.64</td>
</tr>
<tr>
<td>Non-incorporation subgroup (n = 20/14b)</td>
<td>9.94</td>
<td>12.95</td>
<td>8.77</td>
<td>12.14</td>
<td>13.59</td>
<td>12.80</td>
<td>9.78</td>
<td>17.94</td>
</tr>
</tbody>
</table>

Note. IR = implicit recasts; Implicit = implicit knowledge; Explicit = explicit knowledge.
Gain 1 = gain score from pretest to immediate posttest; gain 2 = gain score from pretest to delayed posttest.
a. The incorporation subgroup’s sample size was 6 for implicit knowledge and 3 for explicit knowledge gains.
b. The non-incorporation subgroup’s sample size was 20 for implicit knowledge gains and 14 for explicit knowledge gains.

As Table 72 shows, the sample sizes varied depending on the type of language knowledge. The incorporators gain less language knowledge than the non-incorporators by and large, and, in fact, they showed no improvement in the delayed posttest.

Independent-samples t tests determined there was no significant difference in gain scores between incorporators and non-incorporators (see Table 73). The results indicated that incorporation rate following implicit recasts on 3rd person -s errors was irrelevant to gain scores.

Table 73: T-tests of the IR Incorporators and Non-Incorporators’ Gain Scores for 3rd Person -s

<table>
<thead>
<tr>
<th></th>
<th>df</th>
<th>t</th>
<th>Sig</th>
</tr>
</thead>
<tbody>
<tr>
<td>Implicit knowledge gain 1</td>
<td>24</td>
<td>.39</td>
<td>.70</td>
</tr>
<tr>
<td>Implicit knowledge gain 2</td>
<td>24</td>
<td>-.22</td>
<td>.83</td>
</tr>
<tr>
<td>Explicit knowledge gain 1</td>
<td>15</td>
<td>1.21</td>
<td>.25</td>
</tr>
<tr>
<td>Explicit knowledge gain 2</td>
<td>15</td>
<td>1.25</td>
<td>.23</td>
</tr>
</tbody>
</table>

Note. IR = implicit recasts. *p < .05. **p < .01.

7.4.4 The IR group’s incorporation-acquisition relationship for embedded questions

When implicit recasts were provided on embedded question errors, only two participants produced repair, but none of them incorporated the correct form into the ensuing utterance. As a result, no correlation analysis was carried out for this condition.
7.4.5 A summary of incorporation-acquisition correlations

The correlations between incorporation rate and knowledge gains were found to be statistically significant only when the recasts were explicit and the target structure was salient.

- Incorporation following CR on embedded question errors was correlated with the development of implicit knowledge in the immediate posttest.

- Incorporation following CR on embedded question errors was correlated with the development of implicit knowledge in the delayed posttest.

7.5 Discussion

Question 3 asks whether corrective recasts and implicit recasts lead to different amounts of uptake-with-repair. The answer is affirmative. The chi-square tests report that the extent to which the learners repaired their grammatical errors was significantly related to the type of feedback they received. Following corrective recasts, 21 participants repaired their 3rd person -s errors and 19 participants embedded question errors. Following implicit recasts, seven participants corrected their 3rd person -s errors and two participants embedded question errors. In terms of repair percentage, approximately 64% of CR on 3rd person -s errors and 37% of CR on embedded question errors were repeated or incorporated by the learners while 14% of IR on 3rd person -s errors and 2% of IR on embedded question errors led to repetition or incorporation. In summary, CR resulted in more uptake-with-repair than IR.

By and large, the CR had a higher and the IR a lower repair rate in the current research than recasts in previous studies, for instance, 60% in Loewen and Philp (2006), 18% in Lyster and Ranta (1997), 26% in Mackey and Philp (1998), and 58% in Sheen (2004). All the percentages were based on the total number of provided recasts. These studies did not distinguish explicit from implicit recasts in calculating repair rate, however, a close observation of the examples of the implicit feedback in these studies suggests that it was generally more explicit than the IR in this study. The recasts typically reformulated part of the erroneous utterance and sometimes were as short as three words. Short partial recasts are more explicit (Sheen, 2006). Given the facts that Loewen and Philp’s recasts involved few morphemes, additional stress and segmentation of the target, and that Sheen’s recasts were simple, reduced and stressed, in other words, their CF was overt, it is not unexpected that the repair rates in their studies are close to that after the CR in the current study. The comparison
with previous studies indicates that the explicitness of feedback influences repair rate.

The higher repair rate after corrective recasts than after implicit recasts can be explained by the feedback characteristics. Corrective recasts were made explicit by repeating the learner’s utterance before recasting the error. The repetition serves as an attention getter, which probably draws the learner’s attention (at least, peripheral attention) from meaning to form in the task-based interaction. The stress and rising intonation in repetition helps the learner to focus on the error, and the following recast provides contrastive input. Thus the target structure is more likely to be noticed. Besides, corrective recasts usually repeat and recast only part of an erroneous utterance. These partial recasts may have reduced the cognitive load on learners, and thus were easy to notice. Obviously, corrective recasts are explicit and the explicitness eliminates “the potential ambiguity of recasts” (Lyster, 1998b, p.73). The explicit nature of corrective recasts is reflected in the results of the exit questionnaire. About 91% of the CR participants reported that they noticed the researcher had corrected their grammatical errors in contrast to 76% of the IR participants. The results support Sheen’s (2006) claim that explicit recasts lead to more uptake or repair than implicit recasts.

The higher repair rate after corrective recasts reported in this study compared with that reported for implicit recasts in previous studies can also be explained by the learners’ orientation to form. This study was implemented in a Chinese university, where grammar is valued and learners are more inclined to focus on form. This is evident in the fact that one of the participants in the task control group complained that the researcher did not teach grammar. It stands to reason that the learners’ learning culture primed CR-driven attention to form even though they were participating in meaning-based activities. Furthermore, their orientation to form might have been primed by the fact that they usually had a pair or group discussion to prepare for the task. The importance of participants’ form orientation for uptake-with-repair receives support from Sheen’s (2004) study on the effect of different contexts. Her participants were from Korea and tended to attend to form like other East Asian learners of English, who have been exposed to grammar instruction for years. As in the current research, the teacher in Sheen’s study provided pronunciation and vocabulary practice prior to the observed session. These focus-on-forms sessions might have “helped to foster a general mindset to attend to form” (Ellis et al., 2001, p. 293), and partly contributed to the higher repair rate in response to recasts in her study than in other studies.

The lower repair rate following implicit recasts, particularly implicit recasts on embedded
questions in this study, may be explained by the nature of the target structure in addition to the feedback characteristics. Study 2 targeted embedded questions, recasts on which tended to be longer than those on 3rd person -s because embedded questions involve an inter-clause relationship. The corrective intention in longer recasts is more likely to go unheeded, particularly in the case of implicit recasts, where no attentional strategies prompt learners. Even if perceived, the recasting of an embedded question is apt to be interpreted as another way of saying it. When required to report a question, learners might simply quote the question, believing that it is proper to do so. On hearing the recast, they tend to acknowledge the feedback but make no modification or repair, as illustrated by a participant’s response, “ye ke yi a [what you said will also do]”. This explains why few participants produced repair after implicit recasts on embedded question errors although there were many instances of uptake in the form of acknowledgment. Mackey and Philp’s (1998) low repair rate may also be attributed to their target structure, question forms. Recasts on question formation have to be provided in the form of questions, which sound more like noncorrective repetitions seeking confirmation. As Lyster (1998b) argued, the corrective force of implicit recasts is easily overridden by their functional properties in communicative classrooms.

Question 4 addresses whether there is any relationship between learners’ uptake-with-repair and their subsequent acquisition in terms of implicit or explicit knowledge. In the case of explicit recasts on a salient structure, the answer is perhaps “yes”. Repair rate following corrective recasts on embedded question errors was correlated with gains in implicit knowledge although a negative correlation was found with gains in explicit knowledge. In the case of explicit recasts on a non-salient structure like 3rd person -s, or in the case of implicit recasts, the answer is probably “no”. The correlations in these conditions were either very weak or significantly negative. The correlation analyses using incorporation rate confirmed these claims. Incorporation rate was found to be correlated with the implicit knowledge gains only when embedded question errors were corrected by corrective recasts. By and large, uptake of recasts is not related to subsequent L2 learning.

The finding of negative correlations or no correlation after implicit recasts lends support to Mackey and Philp’s claim that “responses may be red herrings” after recasts (1998, p.338). When implicit recasts are provided, learners’ responses do not indicate their subsequent acquisition of the target structure. The learners might have simply repeated the feedback with little analysis of the linguistic form (Loewen & Philp, 2006) because the corrective force of
implicit recasts is likely to be missed. The corrective intention is further blurred by the meaning-based context. When learners perform a task, “their interest in the task may have overridden their concern for form” (Mackey & Philp, 1998, p.353). Even if they have perceived the corrective intention of the implicit recasts, they may not necessarily succeed in locating the target structure. Even if they have noticed the gap, there is no need for them to repair the error since the correct form has been provided in the recast. On the other hand, as a response to implicit recasts is optional in discourse (Ellis et al., 2001), some participants choose to continue the topic or simply acknowledge the recast. This explains why there were fewer instances of repair after implicit recasts, 11 for 3rd person -s and 2 for embedded questions. Given the extremely low repair frequency, it is premature to make any inference from the negative correlations between repair and gains in explicit knowledge. Nevertheless, it stands to reason that in spite of the lack of repair, the learners might have benefited from the CF as positive evidence, and consequently increased their familiarity with the use and the grammatical rule of the target structure, as illustrated by the immediate or delayed posttest gains. Little repair does not mean that learning does not occur in the implicit recast condition. The efficacy of the feedback should be evaluated by posttest scores rather than repair rate.

Repair following corrective recasts on embedded question errors may be related to L2 learning. This finding corroborates Loewen (2005), and McDonough and Mackey’s (2006) conclusion that repair or target-like modified output predicts subsequent L2 development. However, this is true only when the recasts are provided explicitly on a salient target structure. Corrective recasts were so explicit that their corrective force was easy to perceive. Approximately 91% of the CR participants reported noticing the researcher correcting their errors. A corrective recast immediately juxtaposes the erroneous and correct utterances and adds stress to enhance the corrective force of the feedback. The enhanced corrective force may push learners to produce responses. This is particularly true in a classroom situation. In classroom interaction, learners are inclined to interpret the teacher’s feedback as didactic rather than conversational even though the feedback is provided in meaning-based activities. The explicitness of corrective recasts prompts learners to make such an interpretation, and makes them feel it is necessary to respond linguistically even if their response only consists of a repetition. The “pushed output” (Swain, 1985, 1995, 2000) created opportunities for noticing the gap, and thus potentially facilitated acquisition. Since the language tests and repair rate both gauge the effect of noticing in the case of the corrective recasts on a salient structure, there is a potential relationship between them.
In effect, repair constitutes a form of output that can contribute to L2 learning. The participants may have utilized these output opportunities to proceduralize the encoding and retrieval of the target language knowledge. As mentioned above, the corrective force of corrective recasts is clear, leading to a generally higher repair rate than implicit recasts. The more repair a learner produced, the greater the control of the correct form gained. This explains in part why repair after corrective recasts was significantly related to gain scores for implicit knowledge in the current study.

An interesting finding is that repair rate was negatively correlated with explicit knowledge gains after corrective recasts on embedded question errors. The negative correlation could be due to the complexity of embedded questions and the students’ relatively low proficiency. The formation of embedded questions entails distinguishing main and subordinate clauses and cancelling inversion. This linguistic feature may have been too complex for the learners to work out what the grammatical rule is without the help of explicit instruction. The negative correlation indicates that, far from helping learners develop explicit knowledge, attempts at repair may have confused them. The confusion can be explained by the capacity limits of working memory (WM). WM refers to the mechanism maintaining relevant information during the performance of a cognitive task. In the unitary model and various non-unitary models, there is a wide consensus that the capacity of the WM system or subsystems is limited. The capacity limitation can be conceptualized as the total amount of mental resources available (Engle, Cantor, & Carullo, 1992; Just & Carpenter, 1992). Since embedded questions are a complex target structure, attempts to repair will consume a great amount of cognitive resources. In effect, a simple repetition of the recast embedded question is demanding as it is usually a long utterance. The more resources are used for repairing or repeating, the fewer resources remain for explicit analysis of the structure. Perhaps it was the non-repairers who were more likely to figure out what the complex rule was, hence the negative correlation between repair rate and the development of explicit knowledge for embedded questions.

The second possible explanation for the negative association is the repairers’ low-proficiency. A closer examination of the 15 CR receivers’ repair rate and UGJ total scores revealed that six participants repaired their embedded question errors on more than 50% occasions. Four out of the six participants with a high repair rate scored very low in the UGJ pretest (below 60) and none of them improved significantly from pretest to posttest. As far as the UGJ data
was concerned, it was the low-proficiency learners who repeated or incorporated the correct form more frequently. However, they demonstrated little improvement or even decreased on the immediate posttest, showing confusion about the grammatical rule for embedded questions. Attempts at repair simply confused the low-proficiency learners. In spite of this, repairs after corrective recasts on embedded questions can still facilitate the development of implicit knowledge, as illustrated by the significantly positive correlations between repair rate and implicit knowledge gains. These findings suggest that repair rate following corrective recasts may implicate the acquisition of a salient structure, but not necessarily the development of explicit knowledge, especially when the target structure is complicated.

The last point to be discussed is whether incorporation differs from repetition after recasts as an indicator of L2 development. McDonough and Mackey (2006) claimed that different kinds of responses to recasts are related to language development in different ways. In their study, primed production of the recast form was predictive of subsequent L2 development but simple repetition was not. In the current study, however, neither repair nor incorporation was, by and large, related to L2 learning. The only exception was that both of them were correlated with the development of implicit knowledge in the case of corrective recasts on embedded questions. Repair via repetition or incorporation may indicate acquisition of implicit knowledge in the explicit-recast salient-structure condition.

7.6 Conclusion

The learners in this study produced significantly more instances of uptake-with-repair after corrective than implicit recasts. When explicit recasts target a salient feature, both repair and incorporation were significantly correlated with the development of implicit knowledge. On other occasions, uptake-with-repair following recasts was either negatively or insignificantly related to L2 learning. On the basis of these findings, a number of conclusions are drawn.

- Explicitness in recasts makes a difference in triggering repair, with corrective recasts eliciting more uptake-with-repair than implicit recasts.

- The learners’ orientation to form in the Chinese EFL context partly contributes to the high repair rate after corrective recasts in this study.

- The complexity of embedded questions partly accounts for the lower repair rate after implicit recasts in addition to the implicitness of CF.
• It is likely that repair via repetition or incorporation implicates subsequent L2 acquisition only in the explicit recast condition when the target structure is salient.

• However, by and large, uptake of recasts is irrelevant to L2 learning.

• Repair in response to corrective recasts does not necessarily indicate the development of explicit knowledge.

• Repair after implicit recasts does not confer any benefits.
Chapter 8  Working Memory and Acquisition: Results and Discussion

Chapter 8 explores the relationship between working memory (WM) and recasts-driven L2 learning. The capacity of WM was measured using nonword and listening span tests and the efficacy of recasts was evaluated by means of uptake-with-repair and posttests. The purpose is to examine whether and how the effects of recasts are mediated by different components of WM. The research question to be addressed is as follows.

RQ5: What role does working memory play in corrective feedback involving recasts?

5a. Do differences in learners’ working memory mediate their ability to repair following i) corrective recasts and ii) implicit recasts?

5b. Do differences in learners’ working memory mediate their acquisition of target structures i) in the corrective recast condition and ii) in the implicit recast condition?

8.1 Statistical treatment

There were two working memory tests in the current study: nonword and listening span tests. Nonword repetition is a simple verbal test of phonological short-term memory (PSTM). It required the participants to repeat sequences of nonwords they heard and the maximum sequence length that could be repeated accurately was taken as their PSTM score. The listening span test taps both storage and processing components of WM, which is a mechanism underlying the maintenance of task-related information during the performance of complex cognitive tasks (Baddeley & Hitch, 1974; Daneman & Carpenter, 1980), such as L2 learning. This complex verbal test demanded that the participants should judge the plausibility of each sentence in a set they heard and retain the sentence-end words until they were instructed to recall. A composite Z-score of judgment accuracy, recall accuracy and reaction time was calculated as each participant’s complex working memory score (CWM).

Question 5 addresses the potential link between WM and the recasts-driven L2 learning as reflected by immediate repair rate and gain scores. As the participants’ repair rates were significantly different following corrective and implicit recasts, the WM-repair relationship
was examined in the two conditions separately. Bivariate correlation analyses were conducted for the corrective recast (CR) group in the case of 3rd person -s but independent-samples t tests of the repairers and non-repairers’ WM scores were carried out in the case of embedded questions as the group’s repair data was not linearly distributed as a result of 25% of the participants generating no repair. Similarly, independent-samples t tests were computed for the implicit recast (IR) group in the case of 3rd person -s as the majority of the participants had a repair rate of 0%, but no WM-repair analysis was conducted for the IR group in the case of embedded questions due to the extremely small sample size of the repairers. Only two participants repaired their embedded question errors after receiving implicit recasts.

To reveal the potential mediation of WM on language development, bivariate correlation analyses of WM capacity and pretest-posttest gain scores were performed for the CR and IR groups. In either condition, the correlations were run for two different target structures, 3rd person -s and embedded questions. For both structures, implicit and explicit knowledge scores were used. The former is an average of elicited imitation (EI) and oral production (OP) scores, and the latter is an average of written production (WP) and untimed grammaticality judgment (UGJ) ungrammatical item scores, as justified in Chapter 5 and Chapter 6. Some participants scored over 90 in the WP/UGJ pretest and accordingly were excluded from the explicit knowledge analysis, but they were included in the implicit knowledge analysis due to their lower scores in the EI/OP pretest. As a result, the sample sizes vary according to whether implicit or explicit knowledge scores were entered into the correlation analyses.

8.2 Results for the WM-repair relationship

Bivariate correlation analyses of WM and repair scores and independent-samples t tests of the repairers and non-repairers’ WM scores indicated that the extent to which the participants repaired their errors in response to corrective/implicit recasts was not influenced by their WM capacity (see Appendices O and P for detailed results).

The only exception to the general trend is the significant association between nonword span and repair rate of 3rd person -s errors in the implicit recast condition. An independent-samples t test determined that there was a significant difference in nonword span, \( t(24) = 2.16, p = .04, d = .91 \), between repairers (\( M = 4.09, SD = .77, n = 7 \)) and non-repairers (\( M = 3.44, SD = .65, n = 19 \)) receiving implicit recasts on 3rd person -s errors.
8.3 Results for the WM-gain relationship in the corrective recast condition

Bivariate correlations of WM capacity and gain scores on the measures of implicit/explicit knowledge of 3rd person -s/embedded questions were conducted for the CR and IR groups. The results for the two linguistic features in the CR condition are reported in this section.

8.3.1 Results for 3rd person -s in the corrective recast condition

Descriptive statistics

There were 32 participants in the CR group for 3rd person -s, whose implicit knowledge scores were available for correlation analyses. However, eight of them scored over 90 in the WP/UGJ pretest, and so were excluded from the analyses involving explicit knowledge. The CR group’s sample size for each analysis is reported in Table 74 as well as the means and standard deviations (SD) for WM and gain scores. For each correlation analysis, either nonword or listening span was entered as WM capacity on one hand, and either implicit or explicit knowledge gains were used as L2 learning on the other hand. Therefore, the means for WM and knowledge gains vary depending on different analyses.

Table 74: The CR Group’s WM Capacity and Gain Scores for 3rd Person -s

<table>
<thead>
<tr>
<th></th>
<th>WM</th>
<th>Gain 1</th>
<th>Gain 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
</tr>
<tr>
<td>Nonword span – Implicit (n = 32)</td>
<td>3.46</td>
<td>.56</td>
<td>19.75</td>
</tr>
<tr>
<td>Nonword span – Explicit (n = 24)</td>
<td>3.43</td>
<td>.52</td>
<td>16.62</td>
</tr>
<tr>
<td>Listening span – Implicit (n = 32)</td>
<td>.03</td>
<td>.57</td>
<td>19.75</td>
</tr>
<tr>
<td>Listening span – Explicit (n = 24)</td>
<td>.09</td>
<td>.58</td>
<td>16.62</td>
</tr>
</tbody>
</table>

Note. WM = working memory. Implicit = implicit knowledge; explicit = explicit knowledge.
Gain 1 = gain score from pretest to immediate posttest; Gain 2 = gain score from pretest to delayed posttest.

Correlational analyses

Bivariate correlations of WM and implicit/explicit knowledge gain scores from pretest to immediate/delayed posttest were conducted for the CR group. The results indicate a positive relationship between nonword capacity and implicit knowledge gain for 3rd person -s in both posttests, and that the correlation in the delayed posttest reached a significant level, $r (32) = .36, p = .04$. None of the other correlations was statistically significant (see Table 75).
Table 75: The CR Group’s WM-Gain Correlation Coefficients for 3rd Person -s

<table>
<thead>
<tr>
<th></th>
<th>Implicit (n = 32)</th>
<th>Explicit (n = 24)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gain 1</td>
<td>Gain 2</td>
<td>Gain 1</td>
</tr>
<tr>
<td>Nonword span</td>
<td>.27</td>
<td>.36*</td>
</tr>
<tr>
<td>Listening span</td>
<td>-.25</td>
<td>-.01</td>
</tr>
</tbody>
</table>

Note. *p < .05, **p < .01. Implicit = implicit knowledge; explicit = explicit knowledge.
Gain 1 = gain score from pretest to immediate posttest; Gain 2 = gain score from pretest to delayed posttest.

8.3.2 Results for embedded questions in the corrective recast condition

Descriptive statistics

There were 32 participants in the CR group for embedded questions. All the participants’ implicit knowledge scores were available for analysis, while after deleting two participants with a WP/UGJ pretest score of over 90, 30 participants had explicit knowledge scores available. The CR group’s sample size and means for WM and gain scores in each analysis are presented in Table 76. The means of WM and gain scores vary from analysis to analysis depending on whether nonword or listening span, and implicit or explicit knowledge scores, were entered.

Table 76: The CR Group’s WM Capacity and Gain Scores for Embedded Questions

<table>
<thead>
<tr>
<th></th>
<th>WM</th>
<th>Gain 1</th>
<th>Gain 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
</tr>
<tr>
<td>Nonword span – Implicit (n = 32)</td>
<td>3.47</td>
<td>.57</td>
<td>11.43</td>
</tr>
<tr>
<td>Nonword span – Explicit (n = 30)</td>
<td>3.44</td>
<td>.56</td>
<td>6.58</td>
</tr>
<tr>
<td>Listening span – Implicit (n = 32)</td>
<td>.04</td>
<td>.58</td>
<td>11.43</td>
</tr>
<tr>
<td>Listening span – Explicit (n = 30)</td>
<td>.00</td>
<td>.56</td>
<td>6.58</td>
</tr>
</tbody>
</table>

Note. WM = working memory. Implicit = implicit knowledge; explicit = explicit knowledge.
Gain 1 = gain score from pretest to immediate posttest; Gain 2 = gain score from pretest to delayed posttest.

Correlational analyses

Bivariate correlations of the CR group’s WM and gain scores from pretest to posttest for embedded questions were performed. It was found that listening span was significantly
correlated with implicit knowledge gain in the immediate posttest for embedded questions, \( r(32) = .46, p = .01 \). None of the other correlations was statistically significant.

Table 77: The CR Group’s WM-Gain Correlation Coefficients for Embedded Questions

<table>
<thead>
<tr>
<th></th>
<th>Implicit ((n = 32))</th>
<th>Explicit ((n = 30))</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Gain 1</td>
<td>Gain 2</td>
</tr>
<tr>
<td>Nonword span</td>
<td>.06</td>
<td>-.25</td>
</tr>
<tr>
<td>Listening span</td>
<td>.46**</td>
<td>.03</td>
</tr>
</tbody>
</table>

Note. *\( p < .05 \), **\( p < .01 \). Implicit = implicit knowledge; explicit = explicit knowledge. Gain 1 = gain score from pretest to immediate posttest; Gain 2 = gain score from pretest to delayed posttest.

8.4 Results for the WM-gain relationship in the implicit recast condition

The results for the WM-gain relationship in the IR group are also reported separately for 3rd person -s and embedded questions.

8.4.1 Results for 3rd person -s in the implicit recast condition

Descriptive statistics

Table 78 shows the IR group’s sample sizes, WM means and gain scores for 3rd person -s. There were 33 participants in the IR group. They all had implicit knowledge scores available for analysis, but only 23 of them had explicit knowledge scores available after excluding those who scored higher than 90 in the WP/UGJ pretest. Due to the differences in sample size, WM type and language knowledge type involved in correlation analyses, the IR group’s means for WM and gain scores differ from analysis to analysis.

Table 78: The IR Group’s WM Capacity and Gain Scores for 3rd Person -s

<table>
<thead>
<tr>
<th></th>
<th>WM</th>
<th>Gain 1</th>
<th>Gain 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
</tr>
<tr>
<td>Nonword span – Implicit ((n = 33))</td>
<td>3.59</td>
<td>.70</td>
<td>9.43</td>
</tr>
<tr>
<td>Nonword span – Explicit ((n = 23))</td>
<td>3.48</td>
<td>.66</td>
<td>9.96</td>
</tr>
<tr>
<td>Listening span – Implicit ((n = 33))</td>
<td>-.17</td>
<td>.54</td>
<td>9.43</td>
</tr>
<tr>
<td>Listening span – Explicit ((n = 23))</td>
<td>-.15</td>
<td>.52</td>
<td>9.96</td>
</tr>
</tbody>
</table>

Note. WM = working memory. Implicit = implicit knowledge; explicit = explicit knowledge. Gain 1 = gain score from pretest to immediate posttest; Gain 2 = gain score from pretest to delayed posttest.
**Correlational analyses**

Bivariate correlations of WM scores and gain scores for 3rd person -s were conducted for the IR group. The results show that the correlation coefficients between WM and gains for 3rd person -s in the IR condition were negative, either weakly or moderately (see Table 79). Nonword’s negative correlation was found to be significant with explicit knowledge gain in the immediate posttest, $r (23) = -.41, p = .05$, and marginally significant with explicit knowledge gain in the delayed posttest, $r (23) = -.40, p = .06$.

**Table 79: The IR Group’s WM-Gain Correlation Coefficients for 3rd Person -s**

<table>
<thead>
<tr>
<th></th>
<th>Implicit (n = 33)</th>
<th>Explicit (n = 23)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Gain 1</td>
<td>Gain 2</td>
</tr>
<tr>
<td>Nonword span</td>
<td>-.08</td>
<td>-.04</td>
</tr>
<tr>
<td>Listening span</td>
<td>-.09</td>
<td>-.08</td>
</tr>
</tbody>
</table>

Note. *p < .05, **p < .01. Implicit = implicit knowledge; explicit = explicit knowledge. Gain 1 = gain score from pretest to immediate posttest; Gain 2 = gain score from pretest to delayed posttest.

**8.4.2 Results for embedded questions in the implicit recast condition**

**Descriptive statistics**

The IR group had 33 participants for embedded questions and their implicit knowledge scores were available for correlation analyses. There were 29 participants with explicit knowledge scores as a result of deleting those who scored higher than 90 in the WP/UGJ pretest. Due to the difference in sample size as well as in WM type and language knowledge type, the IR group’s means for WM and gain scores vary according to analyses, as shown in Table 80.
Table 80: The IR Group’s WM Capacity and Gain Scores for Embedded Questions

<table>
<thead>
<tr>
<th>WM</th>
<th>Gain 1</th>
<th>Gain 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>Nonword span – Implicit (n = 33)</td>
<td>3.59</td>
<td>.70</td>
</tr>
<tr>
<td>Nonword span – Explicit (n = 29)</td>
<td>3.57</td>
<td>.67</td>
</tr>
<tr>
<td>Listening span – Implicit (n = 33)</td>
<td>-.17</td>
<td>.54</td>
</tr>
<tr>
<td>Listening span – Explicit (n = 29)</td>
<td>-.12</td>
<td>.55</td>
</tr>
</tbody>
</table>

Note. WM = working memory. Implicit = implicit knowledge; explicit = explicit knowledge.

Gain 1 = gain score from pretest to immediate posttest; Gain 2 = gain score from pretest to delayed posttest.

Correlational analyses

Bivariate correlations of WM and implicit/explicit knowledge gains for embedded questions were performed for the IR group. The results reveal that nonword capacity seemed irrelevant to gain scores for embedded questions, and listening span’s negative correlations with posttest gains were weak in the implicit recast condition (see Table 81).

Table 81: The IR Group’s WM-Gain Correlation Coefficients for Embedded Questions

<table>
<thead>
<tr>
<th>WM</th>
<th>Implicit (n = 33)</th>
<th>Gain 1</th>
<th>Gain 2</th>
<th>Explicit (n = 29)</th>
<th>Gain 1</th>
<th>Gain 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nonword span</td>
<td>-.05</td>
<td>.06</td>
<td>.02</td>
<td>-.03</td>
<td>.03</td>
<td></td>
</tr>
<tr>
<td>Listening span</td>
<td>-.12</td>
<td>-.15</td>
<td>-.19</td>
<td>-.18</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. *p < .05, **p < .01. Implicit = implicit knowledge; explicit = explicit knowledge.

Gain 1 = gain score from pretest to immediate posttest; Gain 2 = gain score from pretest to delayed posttest.

8.5 A summary of the significant WM-acquisition correlations

The extent to which WM mediates explicit/implicit recasts-driven acquisition has been observed at two levels of feedback effect: repair rate and gain scores.

8.5.1 Working memory and repair

In the eight WM-repair conditions (2 WM types x 2 CF types x 2 target structures), only one significant correlation was found:

- Repairers, following implicit recasts on 3rd person -s errors, scored significantly higher than non-repairers on the measure of nonword span.
8.5.2 Working memory and learning

Altogether 32 WM-gain correlation analyses (2 groups x 2 target structures x 2 WM types x 2 knowledge types x 2 posttests) were conducted for the CR and IR groups. Only a few of them were found to be significant.

In the corrective recast condition, there were generally more positive correlations for both target structures. For both structures, one significant correlation was reported with implicit knowledge gains.

- Nonword capacity was significantly correlated with implicit knowledge gains in the delayed posttest for 3rd person -s.
- Listening span was significantly associated with implicit knowledge gains in the immediate posttest for embedded questions.

In the implicit recast condition, the WM-gain correlations tended to be weak and often negative. The only significant correlations were for the explicit knowledge of 3rd person -s.

- Nonword capacity was negatively correlated with explicit knowledge gains, significantly in the immediate posttest and marginally significantly in the delayed posttest for 3rd person -s.

Taken together, the findings differ depending on what feedback was involved and which structure the feedback targeted.

8.6 Discussion

Question 5 was designed to explore the role of WM in corrective feedback. The current study used Baddeley’s (2000, 2007) multiple-component model of WM, which consists of the phonological loop, the visuospatial sketchpad, the central executive and the episodic buffer. With respect to language acquisition, the phonological loop and the central executive subsystems have been thoroughly studied. The phonological loop retains auditory information by articulatory rehearsal, and this phonological short-term memory (PSTM) capacity was measured by a nonword test. The central executive allocates attentional resources by directing, controlling and integrating information from various memory systems,
and retains the processed information in the episodic buffer. The CWM function of simultaneously processing and storing information was captured by a listening span test. The potential mediation of PSTM and CWM on learners’ immediate repair and final language learning was considered separately.

Question 5a asks whether learners’ WM capacity mediates their ability to repair following corrective and implicit recasts. The results of the current study show that repair rate in response to corrective/implicit recasts was not related to WM by and large. Therefore, the learners’ differences in WM were not a major factor in determining whether they repaired their errors after receiving recasts. The only significant association was found between PSTM capacity and repair rate for 3rd person -s in the implicit recast group.

It should be noted that PSTM was only related to repair when implicit recasts targeted the non-salient linguistic feature, 3rd person -s. As Lyster (1998b) argued, implicit recasts are indistinguishable from non-corrective repetitions, and their corrective function is easily overridden by their discourse function in communicative classroom settings. In addition, the redundant non-salient structure increases the difficulty of attending to form. This is evident in the fact that only seven participants produced 11 repairs in the implicit recast group. It stands to reason that those learners with higher PSTM are more likely to perceive the corrective intention of implicit recasts and record the non-salient feature and thus are able to produce “repair” as this repair is in fact just “mimicking” or “repeating” according to Gass (2003, p.236).

Even though those learners with higher PSTM appeared to repair more for 3rd person -s receiving implicit recasts, which was evidence of noticing, this “noticing” did not seem to translate into any learning gains. There was no evidence that those who repaired more learned more in the implicit recast condition. As shown in Chapter 7, the IR repairer subgroup did not significantly differ from the IR non-repairer subgroup in their gain scores for 3rd person -s. In effect, the six repairers gained less than the non-repairers in the posttests. The “red herring” (Mackey & Philp, 1998) nature of uptake of implicit recasts partially accounts for the discrepancy between PSTM’s positive correlation with repair and negative correlation with gains in explicit knowledge, which will be addressed at the end of this section.

In the case of corrective recasts, which are overt, even lower-PSTM learners were able to perceive the corrective intention. Pushed to generate repair, they may perhaps have resorted
to their existing explicit knowledge of the target structures because smaller PSTM capacity tended to fail them in repeating the feedback. The compensatory strategy is particularly true in a Chinese EFL context, where learners have accumulated sufficient explicit knowledge. As the pretest shows, the participants had an explicit knowledge mean score of 70.61 for 3rd person -s and 57.62 for embedded questions. Similarly, learners with low CWM are unlikely to experience difficulty in locating the error because the explicit recasts were usually very brief with stress added to the target structure. Even if they did not identify the structure, the short corrective feedback could be repeated verbatim relying on their PSTM alone and hence led to “successful repair”. That explains why the learners’ repair rate seemed not to be constrained by their PSTM or CWM in the corrective recast condition. This finding sheds light on the current literature on the relationship between WM and successful uptake. Mackey, Adams, Stafford, and Winke (2010) reported that WM capacity was associated with modified output following corrective feedback. The researchers further argued that “the effect of working memory on modified output may be different for different types of modifications” (p. 521) following a post hoc analysis, which showed that WM recall scores were related to repetition-like modification after recasts, but WM processing scores were associated with change-like modification after prompts. From their results, one can infer that PSTM, rather than CWM, contributes to the repetition of recasts. The current research used separate measures of different aspects of WM, finding that although PSTM was potentially related to the uptake of implicit recasts, neither CWM nor PSTM constrained successful uptake when recasts were explicit and short in an EFL classroom.

Question 5b explores whether individual differences in WM mediate CR- and IR-driven acquisition. There were a few significant correlations between WM and implicit/explicit knowledge gains. Only 2 out of 16 correlations were found to be significantly positive for the CR group, and 2 out of 16 correlations were significantly negative for the IR group. In general, the learners’ WM was not a strong mediating factor in their learning of either grammatical structure. This finding is inconsistent with the current literature, which claims that WM is implicated in L2 learning (Miyake & Friedman, 1998; Sagarra & Herschensohn, 2010; Sawyer & Ranta, 2001, just name a few) and in corrective feedback (e.g. N. Ellis, 2005; Mackey et al., 2002; Mackey & Sachs, 2012; Révész, 2012; Robinson, 2005; Sagarra, 2007).

The disparity between this study and previous research can be attributed to the quantity of
corrective feedback. This study was completed in a Chinese university classroom, where there were 32 or 33 participants in each experimental group. Adopting Goffman’s (1981) framework of listeners, we can identify three types of students: addressees, hearers and overhearers. When an individual learner committed a 3rd person -s / embedded question error, the researcher provided corrective or implicit recasts accordingly. The feedback receiver was the one whom the researcher directly addressed while the rest of the class was ratified as hearers. Ideally, “there is little opportunity for learners to function as overhearers” (R. Ellis, 2003, p.39) in task-based language teaching. The hearers were assumed to have attended to the feedback just as the addressee probably did (Kim & Han, 2007), and were consequently included in the correlational analyses. However, in fact, it is unclear what the hearers were doing when the feedback was given. It was possible that some of them positioned themselves as overhearers in such a big class. Even if they positioned themselves as hearers, they were likely to pay limited attention to or not to process the feedback since the feedback was not particularly for them and no response was anticipated from them. In the CR group, eight participants did not make any errors and thus were not corrected. In the IR group, seven participants did not commit 3rd person -s errors and five did not commit embedded question errors, receiving no feedback. Besides, not all the addressees were corrected frequently. For instance, 38% of corrective recast receivers and 58% of implicit recast receivers were corrected only once or twice for their 3rd person -s errors. In other words, the participants were primarily functioning as hearers during most of the class time. It is possible that if each learner had received several instances of feedback, WM would have played a larger role as reported in the previous research. PSTM, CWM or a composite score of both, was found to modulate the short-term or long-term effect of corrective feedback in Mackey et al. (2002), Sagarra (2007), Révész (2012), and Mackey and Sachs (2012). What these studies have in common is that all the learners participated in one-on-one interaction with native speakers of English or computers, during which recasts were provided frequently. “Individualized attention” arguably affects the success of recasts (Han, 2002, p.568), which allows for WM to play a greater role.

The second possible explanation for there being fewer WM-acquisition correlations in this study lies in the quality of corrective feedback. The implicit recast group repaired only two embedded question errors. The extremely low repair rate following implicit recasts on a complex structure suggests that the negative evidence was too implicit for the learners to attend to. If the learners did not process the negative evidence, WM would have a limited role.
to play in the CF. For instance, Li (2013b) and Yilmaz (2013) manipulated recasts as implicit feedback in comparison with explicit correction, reaching a similar conclusion that WM is not implicated in the processing of implicit feedback.

In the corrective recast condition, however, the learners were more likely to perceive the corrective function, notice the target structure and further process it, as this type of CF was quite explicit (Ellis & Sheen, 2006). Of 24 corrective recast receivers, 21 repaired their 3rd person -s errors and 19 repaired their embedded question errors. Benefitting from the explicit CF and ensuing repair, the CR group outperformed the task control group in the EI immediate and UGJ delayed posttests for embedded questions and in the UGJ immediate posttest for 3rd person -s. These results indicate greater processing of the corrective recasts. The processing of the feedback obviously involved WM, which is an attention mechanism. PSTM capacity enables the accurate storage of the feedback and central executive capacity ensures that sufficient attentional resources are available for the deeper processing of the feedback. Learners with greater complex-WM capacity were more able to undertake a cognitive comparison, which resulted in the traces of the correct form forming in the long-term memory. As Yilmaz (2013) noted, “Once discovering the corrective intent stopped being a challenge, learners could have started benefitting from recasts to the extent of their WM” (p. 362). That is why WM was found to be a mediating factor in the case of corrective recasts though the mediating effect was limited in the current study because the participants functioned primarily as hearers (or even overhearers sometimes) rather than addressees.

It is interesting to find that WM mediated the development of implicit knowledge but not explicit knowledge when the recasts were explicit. It can be explained again by the nature of the feedback. According to R. Ellis’s (2006b, p.31) taxonomy of CF, recasts are “input-providing”. Having already provided the correct form, this type of CF does not force or push learners to generate repair by themselves. All they needed to do as a response was to repeat the feedback. They did not have to undertake any conscious analysis. The repetition-like response perhaps only triggered the internal rehearsal in PSTM, which resulted in long-term memory traces as shown in the posttest scores for implicit knowledge. That is to say, the feedback was finally fed into the implicit knowledge system. In order to be fed into the explicit system, the CF has to be deeply processed in CWM. For instance, the feedback needs to be consciously analyzed so as to extract a grammatical rule, and the analysis is then related to whatever explicit knowledge the learner already has in his long-term memory. It can be
speculated that this conscious processing only took place to a very limited extent even in the case of explicit recasts. Similarly, recasts in Révész’s (2012) study resulted in more gains in implicit than explicit knowledge. If this type of CF works better for implicit knowledge, the development of implicit knowledge is more related to WM than explicit knowledge is.

Another interesting finding is that the mediation of WM on language gains in the corrective recast condition varies depending on the target structures. PSTM was significantly correlated with gain scores for 3rd person -s but CWM was for embedded questions. Third person -s is a redundant and non-salient morphosyntactic structure. The acquisition of a non-salient structure relies arguably more on the accurate registration of the feedback in PSTM. The high-PSTM learners were more likely to preserve the recast form through articulatory rehearsal. This advantage in registration, when the corrective intention was obvious, would benefit the locating of the target structure and finally facilitate the development of implicit knowledge. The other target structure, embedded questions, is syntactical, whose acquisition depends more on the mechanism of CWM. Learners need to deploy their central executive to direct attention resources to different components of the clause so that they can perceive the word order. Selective attention needs to be paid to the word order of the clause but inhibition has to be applied to the irrelevant information because the feedback on embedded questions is usually quite long. That is to say, processing embedded questions is cognitively demanding. As Kane and Engle (2000) argued, learners with high complex-WM are more able to resist proactive interference and have more resources available for the processing of attention-demanding target structures. The high complex-WM learners were more likely to succeed in processing embedded questions in the current study.

The interaction between target structures and WM in this study lends support to the view that the processing of different language aspects involves WM subsystems to a varying extent and in different ways (Mackey & Sachs, 2012). The nature of the target structures accounts, at least in part, for the difference in the correlational patterns between WM and implicit knowledge gains in the corrective recast condition.

The last point to be discussed is the paradoxical finding that PSTM was negatively correlated with gains in explicit knowledge of 2nd person -s driven by implicit recasts. It may not be a coincidence that the only positive correlation between working memory and repair was found in the same condition. As discussed earlier, the implicitness of the feedback and the non-salience of the target structure imposed more cognitive demands on PSTM. Greater capacity
in PSTM enhanced the chances of perceiving the corrective intention, and hence showed a strong relationship with repair rate. Due to capacity limits in the whole WM system, on which different WM models agree (Shah & Miyake, 1999), there is a tradeoff between storage and processing. The high-PSTM learners tended to repeat the feedback, the accurate storage of which consumed attentional resources, and so were more likely to have reduced resources available for processing the grammatical rule. In contrast, low-PSTM learners, who produced fewer repair instances, were better able to work out the explicit rule for the target structure. Although speculative, when PSTM was too low to support repetition, learners might not rely on processing the input, but resort to their existing explicit knowledge to engage in interaction. This compensatory strategy may be an explanation for the “less is more” effect of working memory (Miyake & Friedman, 1998; Newport, 1990). Small working memory capacity seems to be beneficial to explicit knowledge for this particular group of learners. As Miyake and Friedman noted, researchers should be prudent when applying this hypothesis to adult learners as their small capacity is actually quite great. The current research tentatively claims that the “less is more” effect may function only in the condition of implicit recasts targeting a non-salient structure when learners already have sufficient explicit knowledge.

8.7 Conclusion
Bivariate correlation and independent-samples $t$ test results reveal that WM capacity does not have much of a role in learners’ ability to repair following corrective/implicit recasts of either structure, with the exception for 3rd person -s in the implicit recast group. Based on the results, a number of speculations are advanced.

- Successful uptake of implicit recasts on a non-salient structure may rely on the capacity of PSTM, which affects the perception of the corrective intention and the accuracy in repeating the CF.

- PSTM’s role in successful uptake of implicit recasts is of little theoretical interest because it does not translate into any long-term effect of CF.

- Working memory (PSTM or CWM) is not a key factor in determining learners’ ability to repair in response to corrective recasts, which are explicit.
Only a few correlations between WM and gain scores were found in this research. The correlation patterns varied according to CF type and target structures. In the corrective recast condition, WM was positively associated with the development of implicit knowledge rather than explicit knowledge, with PSTM for 3rd person -s and CWM for embedded questions. In the IR condition, WM was weakly related to gain scores except for the negative correlations between PSTM and gains in explicit knowledge of 3rd person -s. Conclusions drawn on the basis of these findings are as follows.

- WM seems not to be a decisive factor in the recasts-driven L2 learning in a whole classroom context.

- When recasts are implicit, WM plays a limited role in second language acquisition.

- PSTM might even have a paradoxical effect on explicit knowledge in the condition of implicit recasts of a non-salient feature.

- When recasts are explicit, learners may benefit from the feedback according to their WM capacity.

- WM mediates the acquisition of implicit knowledge in the corrective recast condition.

- WM subsystems are involved in the acquisition of different linguistic features to a varying extent in the corrective recast condition.

- What role WM plays in corrective feedback involving recasts varies depending on the explicitness of feedback, the measure of language knowledge and the nature of the target structure.
Chapter 9  Conclusion

This final chapter reviews the research purpose and the main findings, addresses theoretical and pedagogical implications of the findings, and concludes with limitations of the study and suggestions for future research.

9.1 Purpose of the study

The efficacy of recasts for second language acquisition is highly controversial. It remains unclear whether they constitute an effective technique to provide corrective feedback (CF) in classroom settings and how the CF contributes to L2 acquisition. However, recasts should not be viewed as a single, monolithic corrective feedback strategy. As Sheen (2006) has shown, recasts vary in how implicit or explicit they are. In the current study, two different types of recasts were investigated – simple recasts, which were implicit in nature as the corrective force was not transparent, and corrective recasts, which were much more explicit in nature as they involved a repetition of the erroneous utterance followed, if necessary, by a recast with suprasegmental emphasis. Also controversial is the role of uptake following a recast. Whereas Lyster (1998; 2004) argues that uptake with repair functions as a kind of pushed output that promotes learning, Long (2007) argues that uptake of recasts plays no role in language learning and that the role of recasts is simply to provide learners with positive evidence. The quasi-experimental study reported in this thesis sought to investigate the relative effect of implicit and corrective recasts on both learners’ uptake-with-repair while performing communicative tasks and their acquisition of two target structures as measured by posttests. It also examined the correlation between learners’ uptake-with-repair and their subsequent acquisition.

Recent studies of recasts have shown that the extent to which learners are able to benefit from recasts depends in part at least on their working memory capacity. For instance, Révész (2012) found that high-PSTM (phonological short-term memory) learners benefited more from recasts in oral tests but high-CWM (complex working memory) learners displayed more development in written tests. The study reported in the previous chapters also addressed this controversy by examining the relationship between working memory components (both PSTM and CWM) as measured by the nonword and listening span tests and both uptake-with-repair and acquisition scores.
The focus of both the experimental and correlational studies was on two grammatical structures – 3rd person -s and embedded questions. The former is a non-salient morphosyntactic feature whereas the latter is a complex syntactic structure. Thus the study aimed to also investigate whether the nature of the grammatical structure had any effect on both uptake-with-repair and acquisition and whether the mediating role of working memory differed according to the grammatical structure.

Another important feature of this study was how acquisition was measured. The elicited imitation test (EI) and the oral production test (OP) were oral tests, which were designed to provide a measure of implicit knowledge. The written production test (WP) and the untimed grammaticality judgment test (UGJ) were written tests, which were designed to measure explicit knowledge (R. Ellis, 2005). The validity of these tests as measures of implicit and explicit knowledge was confirmed by a principal-components factor analysis (see Chapter 5). The distinction between implicit and explicit knowledge is important for understanding the role of recasts in L2 acquisition.

9.2 Design of the study

The research reported in this thesis consists of two parts: an experimental study and a correlational study. The design of the experimental study involved four groups with a pretest, treatment, an immediate posttest and a delayed posttest. The tests were designed to provide separate measures of implicit and explicit knowledge. The test control group was not given any treatment, but the corrective recast (CR), implicit recast (IR), and task control groups performed three communicative tasks targeting 3rd person -s and three tasks targeting embedded questions. When committing an error, the learners in the CR and IR groups received corrective feedback accordingly. Thus a comparison of the means between groups using the implicit/explicit knowledge scores was carried out to investigate the relative effect of the corrective and implicit recasts.

The CR and IR groups’ repair rate was calculated and their working memory capacity was measured by means of nonword and listening span tests. Correlation analyses of repair rate, knowledge gain and WM capacity were conducted to investigate whether uptake-with-repair and WM contributed to recasts-driven L2 learning.

9.3 Summary of the main findings

In this section, I will present the main findings for each research question.
RQ1. What effects do corrective recasts and implicit recasts have on learners’ implicit knowledge of L2 grammatical features (3rd person -s and embedded questions)?

1a. What effect do corrective recasts have on learners’ implicit knowledge of L2 grammatical features?

1b. What effect do implicit recasts have on learners’ implicit knowledge of L2 grammatical features?

1c. Is there any difference between the effects of corrective and implicit recasts on learners’ implicit knowledge?

• Those learners who received corrective recasts demonstrated a higher level of implicit knowledge than those learners who received no feedback (i.e. the task control group that only completed the communicative tasks and the test control group that took the tests only). The likely explanation is that the corrective recasts were attended to by the learners because they were relatively explicit. It was proposed that they enabled learners to make the cognitive comparison between their own erroneous output and the feedback, which R. Ellis (1994) has suggested is important for language learning.

• There was no difference between the implicit recast group and the task control group. Implicit recasts did not seem to facilitate the development of implicit knowledge of the two structures. It was likely that the redundant non-salient structure (i.e. 3rd person -s) or the complex structure (i.e. embedded questions) rendered implicit recasts unlikely to work as negative evidence in a communicative setting.

• The corrective recast group outperformed the implicit recast group in the OP delayed posttest for embedded questions. This suggests that the explicitness of recasts is important where acquisition of implicit knowledge is concerned.

• The task control group had higher scores in OP posttest 1 than the test control group for implicit knowledge of the two structures with a large effect size, indicating that oral interaction without any CF is also beneficial for L2 acquisition. The effect of interaction may be attributed to the learners’ form orientation in the EFL context. It is possible that the learners spontaneously focused on form when performing the focused tasks as they were accustomed to treating English as an object of study.
RQ2. What effects do corrective recasts and implicit recasts have on learners’ explicit knowledge of L2 grammatical features (3rd person -s and embedded questions)?

2a. What effect do corrective recasts have on learners’ explicit knowledge of L2 grammatical features?

2b. What effect do implicit recasts have on learners’ explicit knowledge of L2 grammatical features?

2c. Is there any difference between the effects of corrective and implicit recasts on learners’ explicit knowledge?

• The corrective recast group outperformed the task control group in UGJ posttest 1 for 3rd person -s and UGJ posttest 2 for embedded questions. This result indicates that corrective recasts contributed to the development of explicit knowledge. The explicit nature of the CF may have pushed the learners to figure out the grammatical rule for the target structures.

• Those learners who received implicit recasts also had higher scores for 3rd person -s in UGJ posttest 1 and for embedded questions in UGJ Posttest 2 than those who performed the tasks only. This result is surprising but the implicit recasts may have contributed to the development of explicit knowledge due to the learners’ orientation to form in an EFL context. It is possible that these form-oriented learners took advantage of the CF as positive evidence to work on the grammatical rule. Repeated corrections of the grammatical rule in the focused tasks may have helped the development of their explicit knowledge.

• ANCOVAs with UGJ pretest scores for 3rd person -s as the covariate were conducted. When the covariate was set at the level of “mean minus 1 SD”, neither the corrective recast group nor the implicit recast group outperformed the task control group. At the level of “mean”, the corrective recast group outperformed the task control group. At the level of “mean plus 1 SD”, both corrective recast and implicit recast groups did. In other words, the efficacy of corrective and implicit recasts for the development of explicit knowledge varied according to the learners’ initial knowledge of the target structure.
• The corrective recast group had significantly higher scores than the implicit recast group in the UGJ ungrammatical item delayed posttest for 3rd person -s when the covariate was set at “M minus 1 SD” Thus, it can be inferred that explicit recasts were more effective in facilitating the development of explicit knowledge than implicit recasts for learners who demonstrated lower accuracy in the target structure prior to the treatment. This may have been because they were unable to attend to the target structure due to their limited communicative ability.

RQ3. Is there any difference in uptake-with-repair following corrective recasts and implicit recasts?

• Corrective recasts triggered more instances of repair than implicit recasts. This was the case for both structures. The explicitness of corrective recasts may have made the corrective force of the CF more transparent and thus easier to perceive.

• By and large, the corrective recasts resulted in a higher repair rate and the implicit recasts a lower repair rate in this study than the recasts in previous studies. This result confirms previous findings that the explicitness of recasts determines the frequency of immediate repair.

RQ4. Is there any relationship between learners’ uptake-with-repair and their subsequent acquisition of implicit and explicit knowledge?

• Most of the repair-acquisition correlations were not significant and sometimes were negative. Overall, repair as response to corrective and implicit recasts did not implicate subsequent acquisition. One possible explanation for this was that repair following recasts, especially implicit recasts, only involved mimicking the feedback, which did not benefit L2 learning.

• Repair in response to corrective recasts on embedded questions and repair after implicit recasts on 3rd person -s were negatively correlated with the development of
explicit knowledge. The attempts to repair by repeating the feedback may have been too cognitively demanding, preventing the learners from undertaking the analysis needed for explicit knowledge. Those learners who did not repair were more likely to take the opportunity to develop explicit knowledge.

- The only exception was that repair following corrective recasts was positively correlated with implicit knowledge of embedded questions. When the CF was explicit and the target structure was salient, repair, even in the form of mere repetition, was likely to contribute to L2 acquisition.

- Generally speaking, repair with incorporation, like simple repair, was not related to posttest scores in this research.

**RQ5. What role does working memory play in corrective feedback involving recasts?**

5a. Do differences in learners’ working memory mediate their ability to repair following i) corrective recasts and ii) implicit recasts?

5b. Do differences in learners’ working memory mediate their acquisition of target structures i) in the corrective recast condition and ii) in the implicit recast condition?

- By and large, the WM-repair correlations were not significant for either grammatical structure, irrespective of the type of feedback. Neither PSTM nor CWM was related to the learners’ ability to repair. Repair is optional following recasts. Possibly individual difference factors other than WM (e.g. language anxiety, motivation or the extent to which learners are oriented to form) could explain the differences in repair rates of individual learners.

- There was a significant positive correlation between PSTM and repair for 3rd person -s in the implicit recast group. That is, learners with larger short-term memories were better able to repair because the repair is simply a repetition of the CF. However, this repair did not benefit the subsequent acquisition of 3rd person -s, reflecting the fact that repair rates following recasts in general were not related to acquisition (see above).
• Very few WM-gain correlations were significant. Thus by and large, working memory was not a decisive factor in recasts-driven L2 learning in this classroom context. The explanation for this may lie in the fact that learners were exposed to a large number of corrections of both 3rd person -s and embedded questions (see Tables 58 and 59) and as a result were able to process the input provided by the feedback despite limitations in WM.

• In the case of implicit recasts, all the correlations involving WM and gain scores for 3rd person -s were negative. The negative correlations between PSTM and explicit knowledge gains were statistically significant, accounting for 16% of the shared variance. Thus, learners with larger short-term memories generally manifested fewer gains in explicit knowledge. As noted in (2) above, there was a significant positive correlation between PSTM and repair for 3rd person -s in the implicit recasts group. Perhaps the effort that the learners put into repairing their errors following the feedback prevented the analysis needed to develop explicit knowledge.

• In the case of corrective recasts, there were also two significant correlations, both positive and both involving implicit knowledge. Short-term memory was associated with implicit knowledge of 3rd person -s and complex working memory with implicit knowledge of embedded questions. This indicates that different WM components were involved in the acquisition of different linguistic features. A possible explanation for this was that the acquisition of a non-salient morphosyntactic feature relied more on the accurate registration of the CF in PSTM whereas the acquisition of a complex syntactic feature depended more on the distribution of attentional resources in CWM.

RQ6. Is there any difference between the two target structures in uptake rate following explicit/implicit recasts, recasts-driven implicit/explicit knowledge development, and the role of working memory in CF involving recasts?

• For both target structures, corrective recasts led to more instances of repair than implicit recasts. What determines repair rate following recasts is the explicitness of the CF rather than the nature of the target structure.

• By and large, corrective recasts facilitated the development of implicit and explicit
knowledge but implicit recasts enhanced the development of explicit knowledge only. The effect of explicit/implicit recasts on acquisition did not vary according to the target structure.

- Corrective recasts had a superior effect over implicit recasts or tasks alone in the acquisition of embedded questions but not in the acquisition of 3rd person -s. The superior effect of explicit recasts over implicit recasts on the development of implicit knowledge depends on the target structure.

- The superior effect of corrective recasts over implicit recasts on the low-proficiency learners’ explicit knowledge was evident only for 3rd person -s, but not for embedded questions. The latter structure was too complex for the low-proficiency learners to figure out the grammatical rule even if they received corrective recasts, resulting in no difference between explicit and implicit recasts.

- The mediating role of working memory was evident only in the CR-driven development of implicit knowledge, depending on the target structure. PSTM affected the acquisition of a simple non-salient structure while CWM moderated the acquisition of a complex syntactic structure.

9.3 Theoretical implications of the findings

The research reported in this thesis was undertaken within a cognitive-interactionist framework, which posits that corrective feedback (CF) contributes to L2 learning. CF is viewed as an important means for achieving focus on form, enabling form-meaning mappings, and thus facilitating acquisition. The finding that the learners receiving corrective recasts demonstrated greater accuracy in the implicit and explicit knowledge of both target structures in this study provides support for cognitive-interaction theory. The findings of this study also shed light on the following controversies regarding recasts, a specific type of CF.

- Explicitness of recasts;
- Salience of target structures;
- The role of uptake-with-repair;
- The role of working memory capacity;
The role of the instructional context.

9.3.1 Explicitness of recasts

Recasts used to be characterized as implicit corrective feedback (Long & Robinson, 1998; Van den Branden, 1997) and have attracted a lot of research interest. Some researchers advocate that they are an ideal type of CF to facilitate L2 learning, because the correct form this CF provides can be juxtaposed with the learner’s erroneous output, resulting in a cognitive comparison (Doughty, 2001; R. Ellis, 1994). It does not disturb the communication flow so that learners’ cognitive resources can be distributed in an effective way, and attention can be paid to meaning and form simultaneously (Long, 2007). In contrast, Lyster argues that recasts are potentially indistinguishable from noncorrective repetitions or confirmation checks due to their implicitness (1998b) and may go unheeded (2004).

Previous empirical studies reported mixed results for recasts, effective in some studies (e.g. Leeman, 2003; Long et al., 1998) but less so or ineffective in others (e.g. Ammar, 2008; Ammar & Spada, 2006; Ellis, 2007; Lyster, 2004). The limitation of recasts was usually attributed to their implicit nature. Implicit recasts are perceptually non-salient while explicit recasts are more likely to be attended to (Nassaji, 2009), so it is the latter CF that is more likely to facilitate L2 learning according to the Noticing Hypothesis (Schmidt, 1990, 1995, 2001). However, only a few studies have investigated whether explicitness in recasts is important for acquisition. Erlam and Loewen (2010) did not find any difference between the effects of explicit CF (i.e. repetition + recast with declarative intonation) and implicit CF (i.e. single recast with interrogative intonation). In fact, their implicit recasts were more salient with interrogative intonation in the laboratory setting (as discussed in Chapter 5). When explicit and implicit recasts were manipulated in a clear-cut way in classroom-based studies, Taddarth (2010) and Chen (2010) reported that the former was more effective in L2 learning. The current study also found that the explicit recast group outperformed the implicit recast group in the implicit knowledge delayed posttest for embedded questions and in the explicit knowledge delayed posttest for 3rd person -s in a classroom setting. This finding provides support for the claim that explicitness in recasts affects their efficacy. As Ellis and Sheen (2006) suggested, recasts should be viewed as an explicit-implicit continuum.
9.3.2 Salience of target structures
The effectiveness of recasts also depends on the salience of the target structure. Recasts seemed to facilitate the acquisition of salient structures, but not non-salient linguistic features (for instance, Leeman; 2003; Mackey, 2006; Ortega & Long, 1997). However, when perceptual salience was manipulated as an independent variable to test the Salience Hypothesis, Ono and Witzel (2002, as cited in Yilmaz & Yuksel, 2011) found that the results for salience were mixed, with no difference between the salient progressive -ing and the less salient past -ed but with greater development of the salient pronouns than the less salient plural -s. Yang and Lyster (2010) reported that the salient irregular past form benefited more from recasts than the regular past -ed but Yilmaz and Yuksel (2011) showed that there was no difference between salient structures (Turkish plural forms) and non-salient structures (Turkish locative case morphemes). Long (2007) called for further studies to investigate the impact of different target structures on the effect of recasts. The study reported in this thesis targeted 3rd person -s, which was less salient but simpler than embedded questions, which were more salient but more complex. It was found that implicit knowledge of both structures improved after corrective recasts and that explicit knowledge of both structures developed after both corrective and implicit recasts. To put it simply, there was no difference between the recasts-driven acquisition of the two target structures. Target salience did not play a role in recasts in the current study, probably due to the interaction with target complexity. The non-salient 3rd person -s was attended to because it was a simple grammatical rule, as illustrated by the higher repair rate of 3rd person -s errors than embedded questions in response to the corrective and implicit recasts. How different attributes of the target structure affect the efficacy of recasts awaits further investigation.

However, the superior effect of corrective recasts over implicit recasts depended on the target structure. The former was more effective than the latter in the acquisition of embedded questions, which are salient, and in the learning of 3rd person -s, which is simple. The relative effect of explicit recasts on implicit knowledge may be moderated by the salience of the target structure but their benefit for explicit knowledge is probably constrained by the complexity of the target structure.

9.3.3 The role of uptake-with-repair
Recasts have been claimed to be less effective than output-prompting CF as they trigger fewer instances of uptake (for instance, Lyster & Ranta, 1997). According to Swain’s (1985,
Output Hypothesis, CF that pushes greater modified output will contribute more to L2 learning. However, Mackey and Philp (1998) contended that response to recasts is a “red herring”. It does not accurately reflect the cognitive processes in learners’ working memory at the time of receiving the CF. The failure to uptake recasts does not mean they lack efficacy. Long (2007) noted that uptake of recasts should not be equated with language learning, which is a more complex process and takes longer time to occur. Meta-analyses (e.g. Mackey & Goo, 2007; Li, 2010) have provided evidence to show that implicit CF has a bigger effect size in the long run than in the immediate effect. It remains unclear whether uptake plays a role in the feedback of recasts.

McDonough and Mackey (2006) addressed the quality of uptake following recasts. They distinguished different kinds of responses and found that primed production of the recast form was predictive of subsequent L2 development. The current study also explored what form of uptake constituted the pushed modification that promoted language acquisition. It was found that repair via repetition or incorporation was generally irrelevant to the relationship between recasts and language learning. The immediate repair (including incorporation) following recasts may only involve mechanical repetition, which may not entail any noticing of the negative evidence (Gass, 2003). The learners may have simply repeated the feedback with little analysis of the linguistic form (Loewen & Philp, 2006). By and large, repetition and incorporation following recasts did not serve as either an efficient measure of noticing or a good indicator of subsequent language learning. The only positive correlation between repair/incorporation and implicit knowledge appeared in the case of corrective recasts on embedded questions. This finding suggests that when corrective force is transparent, learners’ immediate repair may sometimes assist L2 acquisition. There are conditions for uptake-with-repair to serve as evidence of learning, namely the explicitness of CF and the salience of target structure.

9.3.4 The role of working memory
Interactionist researchers have also posited that individual learner differences mediate the effectiveness of recasts. Whether recasts facilitate language learning depends on the extent to which the learners attend to and process the CF in their working memory (WM), and thus WM, a key component of language aptitude, is implicated in recasts-driven L2 learning (N. Ellis, 2005; Robinson, 2005). To date, however, only a few studies have explored the relationship between WM, recasts and L2 learning, and the results they have reported are
inconsistent. Phonological short-term memory (PSTM) was found to correlate negatively with the immediate effect of recasts in Mackey et al. (2002) but positively in Révész (2012). CWM was a factor in the recasts-acquisition link in Goo (2012) and Sagarra (2007) but not in Li (2013a) and Trofimovich et al. (2007).

Révész (2012) attempted to address this disparity about the role of WM by employing different measures of language knowledge, claiming that PSTM was crucial to proceduralization of emergent knowledge as tapped by oral tests but CWM contributed more to the development of declarative knowledge as tapped by written tests. The current study also administered oral and written tests of language achievement, finding very few correlations between WM and repair/acquisition. Neither PSTM nor CWM constrained the learners’ ability to repair in the EFL classroom, possibly because response to recasts was optional and other cognitive factors or affective factors played a more important role in determining whether they responded to the CF. In such an instructional setting, the learners are exposed to a large number of corrections so that even the low-WM learners may have an opportunity to process target forms. Thus, there were few WM-gain correlations. These findings indicate that the efficacy of recasts is likely to be mediated not by one single individual difference factor but probably by many factors. The interaction between individual difference factors in the recasts-driven language acquisition may form a new avenue of research. The second finding is that both PSTM and CWM were related to the development of implicit knowledge when the recasts were explicit, with PSTM related to 3rd person -s and CWM to embedded questions. This suggests that different WM aspects may be involved in acquisition following recasts, depending on the complexity of the target structure.

9.3.5 The role of the instructional context

Contextual variables also potentially moderate the effectiveness of recasts. Recasts have been reported to be facilitative of L2 learning in laboratory-based studies, whereas classroom research produced mixed results for the role of recasts. For instance, they were facilitative in Goo (2012) but less effective in Lyster (2004). Foreign language contexts lead to a larger effect size for recasts (Li, 2010; Mackey & Goo, 2007) than do second language contexts. Based on Lyster and Mori’s (2006) Counterbalance Hypothesis, recasts as an implicit form of CF work well in EFL settings because the CF counterbalances learners’ predominant orientation to form. Sheen (2004) examined the role of context in corrective feedback using uptake as a measure. She attributed the greater instances of repair in Korean EFL and New
Zealand ESL classrooms than in Canadian ESL and immersion classrooms to the learners’ orientation to form in the former two contexts as well as the explicit characteristics of the recasts. Up to now, however, no empirical study has investigated how context mediates recasts-driven development of language knowledge.

The current study was not designed to compare different contexts, but the finding that corrective recasts worked for both implicit and explicit knowledge and implicit recasts worked for explicit knowledge in the Chinese EFL setting is indicative of a role for context. In the Chinese EFL classroom, learners are generally form-oriented. This can explain why the task control group outperformed the test control group in implicit knowledge gains even though it received no feedback. This finding lends support to the Counterbalance Hypothesis.

**9.4 Pedagogical implications of the findings**

In this section, main findings from the exit questionnaire are briefly summarized as they are important for understanding pedagogical implications of the task-based materials.

- Most participants (91%) reported noticing of the corrective recasts and about half of them correctly located the errors (67% for 3rd person -s and 42% for embedded questions). The corrective recasts tended to be noticed successfully.

- Three quarters of the participants (76%) claimed that they perceived the corrective force of implicit recasts but only a few of them recalled the target structure (21% for 3rd person -s and 9% for embedded questions). The focus of implicit recasts was seldom noticed successfully.

- Most participants (92%) were open to corrective feedback, stating that corrective feedback was useful for language learning.

- The majority of task participants (80%) had a positive attitude towards the task-based project, some commenting that the tasks were interesting and interactive, some remarking that this type of instruction provided more opportunities for speaking, some positing that the task-based approach was better than the traditional teaching method, and even one suggesting that this project should be integrated into the normal learning program.
The findings regarding the effect of tasks and corrective feedback and the relationship between uptake-with-repair, language learning and working memory in the current study allow room for a number of pedagogical implications.

9.4.1 Modular syllabus
The participants showed great interest in the task-based approach, which was new to them. Their regular classes are, by and large, form-focused instruction, constrained by English textbooks, teachers’ experience and beliefs, and the difficulty in implementing meaning-focused instruction in China, among other factors. The tasks utilized in the current study were attractive to the learners, as reflected by their nearly full attendance, active participation in task performance and very positive comments on the whole project in the exit questionnaire. They were strongly motivated to learn English through tasks. Therefore, EFL teachers who prefer form-focused instruction are advised to design and incorporate communicative tasks when appropriate to promote learners’ motivation and English proficiency. The “communicative opportunities” (Ellis & Shintani, 2014, p. 48) provided by the tasks can serve as supplements to traditional instruction. At least, a modular syllabus may be most appropriate for the Chinese EFL classroom context.

9.4.2 Focused tasks
It is advisable to draw on focused tasks in an EFL classroom. Focused tasks “elicit the use of language features that learners might otherwise not attend to and avoid using” (Ellis & Shintani, 2014, p.138). Attention to a large number of exemplars of a specific predetermined feature assists form-meaning mappings. The effect of tasks is evident in the fact that the task control group outperformed the test control group in the development of implicit and explicit knowledge. The task effect in this study suggests the feasibility of implementing focused tasks in an EFL classroom context where learners are accustomed to making language the object of study, although Littlewood (2007) doubted it. As Ellis and Shintani argued, task-based language teaching constitutes an “ideal” approach in foreign language contexts because the learners have limited opportunity for communication outside classrooms (2014, p. 156). Thus, Chinese EFL teachers can conduct analyses to find out what linguistic features are difficult for learners to acquire. Focused tasks targeting these problematic features will foster L2 learning.
9.4.3 Explicit recasts and multi-move feedback
Corrective recasts were more effective than implicit recasts in the development of implicit knowledge for a salient structure (i.e. embedded questions) and in the development of explicit knowledge involving those learners who had lower prior knowledge of a simple structure (i.e. 3rd person -s). This finding indicates the superiority of explicit recasts (i.e. repetition + recast + stress) over implicit recasts in L2 learning. It is suggested that EFL teachers should use emphatic stress on a non-salient structure to achieve “noticing instruction” (Ellis & Shintani, 2014, p.185). The stress may attract attention to form. Given the advantage of multi-move recasts (Sheen, 2006), that is, corrective recasts in this study, it may be beneficial for EFL teachers to combine recasts with other types of feedback as long as the CF does not unduly interrupt the communicative flow. A combination of output-prompting feedback and the input-providing recasts may be particularly effective as the former elicits a response, potentially benefiting increasing control of an existing structure, and the latter presents positive evidence, benefiting acquisition of a new structure (Ellis & Shintani, 2014).

9.4.4 Immediacy of corrective feedback
The effectiveness of CF (i.e. corrective recasts and implicit recasts) provided while the learners were performing a task suggests that it is not necessary to delay correction, which Gattegno (1972) and Harmer (1983) advocated in order not to interrupt communication. Neither corrective recasts nor implicit recasts were intrusive, and both of them constituted “brief, unobtrusive immediate correction” (Scrivener, 2005). Most negotiation of form episodes were four to six turns long and did not interfere with communicative fluency in the current study. Ideally, EFL teachers should provide recasts immediately after hearing a learner’s erroneous utterance so that the juxtaposition of correct feedback and erroneous utterance affords an opportunity for cognitive comparison.

9.4.5 Irrelevance of uptake after recasts
Uptake after recasts proves to be irrelevant to subsequent L2 learning in the current study. As Long claimed, output is not a necessary condition for L2 learning although potentially facilitative (1996), and uptake of recasts is not necessary (2007). Even if the learners do not produce uptake-with-repair, it does not mean that they are not processing the feedback. What is crucial to L2 learning is not uptake-with-repair following recasts, but learners’ noticing of the target structure (Ellis & Mifka-Profozic, 2013). There is no need for EFL teachers to push or urge the learners to produce immediate repair after recasts.
9.4.6 Relationship between individual differences and task types

Working memory was found not to be a decisive factor in the recasts-driven language learning, probably due to the intensity of the feedback directed at a specific feature. The greater opportunities for attending to the target feature the focused tasks created may have minimized the mediation of individual differences factors. That is why focused tasks are an ideal form of instruction. However, in unfocused instruction, L2 teachers probably need to take into account learner individual differences. As Ellis and Shintani (2014) suggested, language teachers are expected to deploy a mix of corrective feedback strategies to accommodate individual differences when using unfocused tasks.

9.5 Limitations of the study

This study was classroom-based. Like all other classroom studies, there were limitations to the design of the study although the payoff was high ecological validity.

The instruction tasks varied in the number of obligatory contexts that were created and the time taken to complete each task. For instance, the “My Partner” task triggered more exemplars of 3rd person -s than other tasks, and the “Jackie Chan” task took more time to complete than other tasks for embedded questions. These differences between tasks may have possibly affected the results. However, it would have been demotivating to use the identical type of tasks for a specific feature for three consecutive sessions. Things seldom happen this way in real classrooms.

The researcher worked as the teacher in this study, which may have influenced the effectiveness of the corrective feedback. This raises a possible limitation in terms of the pedagogical implications of the study. It is not clear whether the same effect of corrective or implicit recasts will be found with the class’s normal teachers. Nevertheless, such a researcher-as-teacher arrangement was necessary for carrying out the treatment, given the scale of the whole project. The researcher designed the tasks and had experience in providing the required feedback in a consistent manner. He knew the instructional materials better than anybody else.

There was no actual evidence of whether noticing took place during the lessons. On the basis of repair rate after corrective and implicit recasts and the results from an exit questionnaire, it was only possible to speculate that the feedback contributed to L2 learning via noticing and
noticing the gap. In the context of the whole study, it was not possible to investigate noticing by means of stimulated recall.

Chinese EFL learners tend to learn English by rote. Some participants were seen trying to memorize the complete text while they were preparing for oral reproduction. They were advised not to do so but their attempts to memorize may have affected the extent to which the oral production test measured spontaneous use of the target structure. To ensure the test validity, time limits were set for preparation, 15 minutes for the pretest, eight minutes for the immediate posttest and five minutes for the delayed posttest, but this may not have completely prevented attempts at rote memorization.

The oral production test was probably too difficult for a few students, who lacked proficiency to complete the task, producing fewer than five obligatory contexts for the target structure. These participants were removed from data analysis. The effect of the treatment on these learners with extremely low proficiency was not investigated and the exclusion of these participants may have affected the results. However, this is acceptable given the large sample size and the complexity of the current study.

There was some evidence of a practice effect in the test results. The test control group scored much higher for embedded questions in the oral production delayed posttest than before although the improvement did not reach a significant level. It was possible that some of the treatment effect may have resulted from the test practice. This limitation could have been avoided if there had been three different tests, which were counterbalanced. The counterbalanced design was not adopted however because it would have caused a lot of logistical problems in this classroom-based study. The teacher would have had to divide the class into groups and give them different tests, which was not possible given the complexity of the study and the participants’ timetables. However, the practice effect was not a major issue because the test control group served as a basis for group comparisons to check the effects of corrective and implicit recasts and tasks.

Working memory was measured by means of nonword span, digit span and listening span tests, but the digit span test was subject to “cheating”. About 20% of the participants deliberately or unconsciously jotted down the numbers they heard, affecting the validity of the test as a measure of phonological short-term memory. As a result, the results from this test
were excluded from analysis. The learners’ phonological short-term memory of meaningful items was not investigated.

9.6 Suggestions for future research

All the participants in the CF groups had an opportunity to complete a task directed at the target feature but those who had not committed an error in the target structure did not receive any feedback. It was not observed whether they processed and benefited from the feedback that was provided to other learners. It will be interesting to see whether CF works differently for addressees and hearers.

This study only used uptake-with-repair following the feedback and self-report in the exit questionnaire as indicators of noticing. Noticing was not tapped by any direct means. Future research on the relationship between recast types, noticing and L2 learning may draw on stimulated recall protocols or some other online self-report mechanism when time permits. A direct test of noticing will also assist in clarifying the link between noticing and repair after recasts.

The participants’ oral reproduction was analyzed by means of target-like use analysis, which reported accuracy of the target structure against the total of obligatory and non-obligatory contexts. It is still worthwhile to focus on obligatory contexts to ascertain whether or not the number of obligatory contexts each participant creates will increase after receiving recasts.

The effect of corrective or implicit recasts did not differ between the salient (i.e. embedded questions) and non-salient structures (3rd person -s). This was attributed to the interaction between salience and complexity of the target structures. This hypothesis regarding target structures needs further empirical support. The participants had already gained certain knowledge of both structures prior to the treatment. Thus, corrective recasts promoted the development of implicit knowledge, but implicit recasts did not. The value of implicit recasts may lie more in the internalization of a new form than in the increasing control of an old form. New and old linguistic forms can be compared in the future to see what role recasts play in the acquisition of both.

As discussed in the theoretical implications section, the efficacy of recasts may be simultaneously mediated by many factors. For instance, learners with low working memory may resort to their language analytic ability to produce repair as they have prior knowledge
of the target structure and thus may benefit from the feedback. The interaction between individual difference factors in recasts-driven language acquisition awaits future research.

Notes.

1. Loewen and Nabei (2007) completed their study in an EFL classroom setting, but organized the treatment class into groups of four students, and the groups separately received different types of CF from one of the three researchers. This kind of quasi-experimental design resembled that of laboratory studies.
2. Examples of repetition and incorporation following corrective recasts are provided as below.
   
   S: Yes. And he often make a laugh.
   T: He often.
   S: make a laugh.
   T: He often MAKE? He often makeS.
   S: Makes. [repetition]
   
   S: She come from Hunan, Hunan.
   T: She COME?
   S: Come, come.
   T: She comeS.
   S: Comes. She comes from Hunan TV Talents Show. [incorporation]

3. Three working memory tests were actually administered for this study, namely nonword, digit and listening span tests. Unfortunately, the digit span data were rendered invalid because about 20% of the participants “cheated” deliberately or unconsciously. After all, numbers are subject to jotting down. Consequently, only nonword span and listening span scores were entered into statistical analyses.
Appendix A  Participant Information Sheet (English Version)

PARTICIPANT INFORMATION SHEET
(For Participants)

Project Title: The effect of recasts, noticing and working memory on L2 acquisition
Name of Researcher: Yongbin Zhao

Researcher Introduction
I am Yongbin Zhao, a PhD student in the Department of Applied Language Studies and Linguistics, the University of Auckland. My student ID is 2628651.

Project Description and Invitation
The project involves the students participating in task-based lessons. The focus of my research is on the classroom interactions that arise in these lessons. I will analyze the specific aspects of the naturally-occurring interactions between students and teachers. Such an analysis will shed light on the process of second language teaching and learning.

I am going to do this project in your university by offering task-based lessons in your spare time. You will be taught over 6 weeks and each week you are supposed to come to a 1-hour session. I will organize different types of language learning tasks, such as story telling and answer guessing, passage reading, interviewing and reporting, etc. You can take these opportunities to practice your oral English. I am sure you can enjoy the fun of learning English here!

Project Procedures
1. Project duration: about 20 hours over 11 weeks (February - June 2012)
2. Schedule:
   Week 1 / 5: All groups take the pretest (four knowledge tests)
   Week 2-4 / 6-8: Instruction groups receive task-based instruction (one session per week)
   Week 4 / 8: All groups take Posttest 1 (four knowledge tests)
   Week 9 / 13: All groups take Posttest 2 (four knowledge tests)
   Week 10: All groups take memory capacity tests
   Week 13: Instruction groups complete an exit questionnaire.
3. Benefits and Risks
   The potential risk is that you might become a little bit busier in your spare time as participation takes 1 hour a week. However, the practice will be beneficial to your language learning.
   During the 6-week instruction, you can learn English through a different teaching approach – task-based language teaching. It gives you more opportunities to practice your oral English and to improve your English grammar.
4. Compensation
As compensation, 10 NZ dollars’ worth of gifts will be presented to all participants. Besides, I will help you to learn English by answering your questions for 3 office hours per week.

At the completion of the project, follow-up classes will be offered for the control group to obtain the instruction that other groups have received. It is your right to accept or give up these opportunities.

5. Funding for the research

Funding for this research is being sought from Faculty of Arts, the University of Auckland, where I am enrolled. Any further information will be provided for your reference.

6. Participation / nonparticipation

I have got your name list from the university authority, which enables me to come and invite you to take part in the project. However, participation is voluntary and you have the right to refuse or agree. The principal has agreed that neither your refusal nor your agreement will affect your grades and neither will affect your relationship with the university.

Data Storage / Retention / Destruction / Future Use

1. Data collection

Interactive sessions and oral tests will be audio-recorded and written tests will be administered in the form of questionnaires. Consequently, there are recorded data, hard-copy data and electronic data.

2. Data storage

The data will be kept for 6 years. The tapes will be stored separately from the transcript and other identifiable information, and only my supervisor and I can get access to them. Hard copies will be locked in University premises under my own control. Electronic data will be stored only in the hard-disk drive of my personal computer, which are not allowed to be copied after a backup has been made. The electronic files and their backup are accessible only by my supervisor and me through the use of codenames.

3. Data use

The data are collected for my PhD thesis, conference presentations and future publications. They will not be used for other purposes.

Questionnaires and transcripts will be photocopied with your identifiable information deleted and will be numbered 1-499 before being scored by another rater, who has signed a rater confidentiality agreement. The list linking you with the questionnaire will be maintained only by the researcher.

All data will be reported and published in a way that does not disclose your identifiable information so that your confidentiality is preserved.

4. Data destruction

At the completion of the research and its publication, the tapes will be cut into pieces, the hard copies will be burned out, and the electronic data will be completely deleted by formatting the disk.

Right to Withdraw from Participation

As participants, you are free to withdraw from the research at any time. You also have the right to withdraw your hard-copy data and removable separate electronic data on the condition that you have informed me of your intention prior to the data analysis (by 1 June 2012). The university authority has agreed that your withdrawal of participation and withdrawal of information will not affect your grades and your relationship with the university. What should be noted here, however, is that you are not allowed to withdraw any recorded data. It is almost impossible to cut apart your recording without destroying the completeness of your conversation with another interlocutor.

Anonymity and Confidentiality

It is impossible to keep you anonymous during the 6-week instruction. However, your anonymity and confidentiality are guaranteed in any record of your participation (see the section of Data Storage / Retention / Destruction / Future Use).

Research Findings

When the results have been worked out in a year or so, I will email your principal, and he will help me to inform you of the research findings.

Contact Details and Approval Wording

Thank you for reading this information sheet, and if it is possible, participating this project. If you have
any questions, please contact me.

Yongbin Zhao (PhD)
Department of Applied Language Studies and Linguistics
The University of Auckland
Private Bag 92019
Auckland 1142
New Zealand
Email: yzha586@aucklanduni.ac.nz (New Zealand)
       zhaomaya@yahoo.cn (China)
Phone: +64 210761335 (New Zealand)
       13315181615 (China)

My supervisor is Professor Rod Ellis. His contact details are:
Professor Rod Ellis
Department of Applied Language Studies and Linguistics
The University of Auckland
Private Bag 92019
Auckland 1142
New Zealand
Email: r.ellis@auckland.ac.nz
Phone: 3737599 (Ext. 84876)

The Head of Department is Yan Huang. His contact details are:
Professor Yan Huang
Department of Applied Language Studies and Linguistics
The University of Auckland
Private Bag 92019
Auckland 1142
New Zealand
Email: yan.huang@auckland.ac.nz
Phone: 3737599 (Ext. 87809)

For any queries regarding ethical concerns you may contact:
The Chair, The University of Auckland Human Participants Ethics Committee, The University of
Auckland, Office of the Vice Chancellor, Private Bag 92019, Auckland 1142. Telephone 09 373-7599
extn. 83711.

APPROVED BY THE UNIVERSITY OF AUCKLAND HUMAN PARTICIPANTS ETHICS
COMMITTEE ON 11 May 2011 for (3) years, Reference Number 2011 / 209.
致参与者信息单
（致学生）

项目名称：任务型教学中的重铸反馈、注意与工作记忆

研究者姓名：赵永彬

研究者简介
赵永彬，奥克兰大学文学院应用语言研究及语言学系在读博士生，学号 2628651。

项目概述及邀请
本项目调查的是任务型教学中学生的参与情况，研究的焦点是教学中自然产生的师生互动，主要分析师生互动的各个方面。这些分析将进一步深化我们对第二语言教学过程的理解和认识。

我想在大家的业余时间开设英语口语课，在口语课上实施任务型教学，以此来完成项目调查。整个项目需要 11 个星期，其中口语课持续 6 周，每周一次课，每次 1 小时。课上我们将组织各种英语学习活动，包括讲故事、猜答案、读文章、采访同学、总结汇报等语言使用任务。大家可充分利用这些机会练习英语口语，在轻松愉快的环境中提高英语交际能力。来吧，英语学习原来也是快乐的！

项目程序
1. 项目时间：整个项目共 20 个小时左右，分 11 周完成（2012 年 2 月至 6 月）
2. 研究进程：
   第一 / 五周：前测（英语知识测验）
   第二至四 / 六至八周：任务型教学（每周一次，每次 1 小时）
   第四 / 八周：后测 1（英语知识测验）
   第九 / 十三周：后测 2（英语知识测验）
   第十周：记忆力测试
   第十三周：结束问卷
3. 利弊
   因每周每名参与者需要 1 个小时左右的时间，您的业余时间可能会稍微忙碌一些。
   但参与是值得的，在为期 6 周的教学中，您将体验到一种全新的英语教学模式，即任务型教学法。在这种教学过程中，您有更多口语练习的机会，迅速提高英语交际表达能力，英语语法也随之增长。
4. Compensations

During the experiment, I will work an additional 3 hours each week for our school, focusing on answering questions and helping everyone learn English. At the end of the project, each participant will receive a small gift worth approximately 10 NZD as a token of appreciation.

5. Research Costs

Researchers have already approached the university, i.e., the University of Auckland Faculty of Humanities, to apply for doctoral research funds. If approved, they will inform everyone.

6. Participation/Non-participation

I have already provided the list of participants, but we will take voluntary participation. You have the right to participate or not. The department head has agreed that participation or non-participation will not affect your academic assessment or relationship with the school.

Data Storage, Retention, Destruction, and Future Use

1. Data Collection

Experiments and oral tests will be recorded, and written tests will be taken in the form of questionnaires. The results consist of three types of data: audio, written, and electronic data.

2. Data Storage

Data storage period is 6 years. Audio and audio text are saved separately, and only I and my supervisor can use them. Written materials are stored in my office at Auckland University, and only I can access them. Electronic data is stored on my personal computer, and all related electronic documents and their backups can only be opened and used by me and my supervisor with a password.

3. Data Use

Data only allows researchers to conduct research, participate in academic conferences, and use data for future publications. No other use is allowed. Copies of questionnaires will be removed before the second evaluator (I am the first evaluator). Questionnaire编号 and participant personal information are only used by me. The second evaluator has signed a confidentiality agreement and maintains strict confidentiality.

All data reports and publications will not reveal participants' identities, and your personal information will be strictly protected.

4. Data Destruction

After research publication, audio tapes will be destroyed, written materials will be burned, and electronic documents will be deleted using magnetic formatting.

Withdrawal Rights

Participants can withdraw at any time during the project. If you withdraw your paper and independent electronic data before data analysis begins (2012/6/1), you also have the right to withdraw your information. The department head has agreed that withdrawal of information or above will not affect your academic assessment or relationship with the school.

Anonymous and Confidentiality

Excluding the first 6 weeks, participants can remain anonymous. Any information and data related to your participation will be protected (see “Data Storage, Retention, Destruction, and Future Use” section).

Research Results

Research results will be sent in email to the department head. The department head will inform everyone.

Contact Information

Thank you for reading the above information! If you agree to participate in this project, please sign the consent form. Thank you very much! If you have any questions about the project, please contact me.

Email: Yongbin Zhao (PhD)
Department of Applied Language Studies and Linguistics
The University of Auckland
Private Bag 92019
Auckland 1142
New Zealand
电子邮件: yzha586@aucklanduni.ac.nz (新西兰)
zhaomaya@yahoo.cn (中国)

电话: +64 210761335 (新西兰)
13315181615 (中国)

或联系我的导师 Rod Ellis 教授:
Professor Rod Ellis
Department of Applied Language Studies and Linguistics
The University of Auckland
Private Bag 92019
Auckland 1142
New Zealand
电子邮件: r.ellis@auckland.ac.nz
电话: 3737599 (分机 84876)

或联系系主任 Yan Huang 教授:
Professor Yan Huang
Department of Applied Language Studies and Linguistics
The University of Auckland
Private Bag 92019
Auckland 1142
New Zealand
电子邮件: yan.huang@auckland.ac.nz
电话: 3737599 (分机 87809)

如有任何人类相关研究伦理道德方面的疑问，请与奥克兰大学人类相关研究伦理委员会主席联系:
The Chair, The University of Auckland Human Participants Ethics Committee, The University of Auckland, Office of the Vice Chancellor, Private Bag 92019, Auckland 1142。电话 09 373-7599 分机 83711。

本项目由奥克兰大学人类相关研究伦理委员会于 2011 年 5 月 11 日批准，有效期 3 年，编号 2011 / 209。
### Appendix C  Transcription Conventions

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>teacher</td>
</tr>
<tr>
<td>S</td>
<td>student as an interlocutor</td>
</tr>
<tr>
<td>Other (s)</td>
<td>student(s) other than the interlocutor</td>
</tr>
<tr>
<td>CAPITAL</td>
<td>speaker emphasis / louder voice</td>
</tr>
<tr>
<td>Italic</td>
<td>Chinese pinyin</td>
</tr>
<tr>
<td>[ ]</td>
<td>translation of the transcriber</td>
</tr>
<tr>
<td>,</td>
<td>pause, steady or rising tone of voice</td>
</tr>
<tr>
<td>.</td>
<td>pause, falling tone of voice</td>
</tr>
<tr>
<td>“ ”</td>
<td>quotation</td>
</tr>
<tr>
<td>?</td>
<td>utterance in the form of a question</td>
</tr>
<tr>
<td>-</td>
<td>no response</td>
</tr>
<tr>
<td>=</td>
<td>overlapping</td>
</tr>
<tr>
<td>xxx</td>
<td>inaudible sequence</td>
</tr>
</tbody>
</table>
Appendix D  Elicited Imitation Pre/Post/Delayed Post-test

Answer Sheet

Name: ________  Group: ________.

Directions: Now you are going to hear 18 statements, which you must agree with, disagree with or indicate you are not sure about. So when you hear each statement, first decide whether you agree or not by marking A (true), B (not true) or C (not sure) on the answer sheet, and then repeat the sentence in CORRECT English. Remember to make a decision before you repeat to the voice recorder.

<table>
<thead>
<tr>
<th></th>
<th>A. true</th>
<th>B. not true</th>
<th>C. not sure</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>A. true</td>
<td>B. not true</td>
<td>C. not sure</td>
</tr>
<tr>
<td>2.</td>
<td>A. true</td>
<td>B. not true</td>
<td>C. not sure</td>
</tr>
<tr>
<td>3.</td>
<td>A. true</td>
<td>B. not true</td>
<td>C. not sure</td>
</tr>
<tr>
<td>4.</td>
<td>A. true</td>
<td>B. not true</td>
<td>C. not sure</td>
</tr>
<tr>
<td>5.</td>
<td>A. true</td>
<td>B. not true</td>
<td>C. not sure</td>
</tr>
<tr>
<td>6.</td>
<td>A. true</td>
<td>B. not true</td>
<td>C. not sure</td>
</tr>
<tr>
<td>7.</td>
<td>A. true</td>
<td>B. not true</td>
<td>C. not sure</td>
</tr>
<tr>
<td>8.</td>
<td>A. true</td>
<td>B. not true</td>
<td>C. not sure</td>
</tr>
<tr>
<td>9.</td>
<td>A. true</td>
<td>B. not true</td>
<td>C. not sure</td>
</tr>
<tr>
<td>10.</td>
<td>A. true</td>
<td>B. not true</td>
<td>C. not sure</td>
</tr>
<tr>
<td>11.</td>
<td>A. true</td>
<td>B. not true</td>
<td>C. not sure</td>
</tr>
<tr>
<td>12.</td>
<td>A. true</td>
<td>B. not true</td>
<td>C. not sure</td>
</tr>
<tr>
<td>13.</td>
<td>A. true</td>
<td>B. not true</td>
<td>C. not sure</td>
</tr>
<tr>
<td>14.</td>
<td>A. true</td>
<td>B. not true</td>
<td>C. not sure</td>
</tr>
<tr>
<td>15.</td>
<td>A. true</td>
<td>B. not true</td>
<td>C. not sure</td>
</tr>
<tr>
<td>16.</td>
<td>A. true</td>
<td>B. not true</td>
<td>C. not sure</td>
</tr>
<tr>
<td>17.</td>
<td>A. true</td>
<td>B. not true</td>
<td>C. not sure</td>
</tr>
<tr>
<td>18.</td>
<td>A. true</td>
<td>B. not true</td>
<td>C. not sure</td>
</tr>
</tbody>
</table>
Appendix E  Oral Production Pre/Post/Delayed Post-test

**Directions:** Now we are going to read a passage. Please read carefully because you will be required to reproduce it in details later. You are allowed to make notes of the main points but do NOT write out full sentences. Besides, your notes will be collected prior to the retelling. If you come across any unfamiliar words, you can ask the teacher for help. When you are ready to retell it, please let the teacher know.

**A Typical Day in President Obama’s Life (Study 1)**

President Obama usually gets up early in the morning. He goes for a workout (健身) at 6:45 a.m. He likes to do this before doing anything else because that way he can be sure that he has the time for a workout every day. After that, he eats breakfast with his family and then helps get his children ready for school before walking downstairs to work.

President Obama usually shows up (露面) at the Oval Office (总统办公室) shortly before nine o’clock in the morning. On a typical day, he meets with his advisors and foreign leaders, signs bills and executive (行政的) orders, works on the federal budget (联邦预算), and makes public appearances. He works until about 6:30 in the evening.

President Obama spends every evening with his family. He eats dinner with his family, hangs out (闲逛) with the kids and puts them to bed at about 8:30 p.m. And then he returns to work.

President Obama reads more briefing papers (简报), does a lot of paperwork, and sometimes writes until nearly midnight. He knows that he has more mental energy at night. While much of the country is sleeping, the president thinks more clearly and makes his best decisions. He also dedicates (致力于) some of his time to getting ahead (进步) — learning new things and improving himself. He expands (拓展) his mind every night by reading, not by watching TV. It is interesting that he does not read about politics, foreign affairs or anything related to his “job”. He reads for pleasure. For example, he is currently reading *Netherland* (小说《荷兰》), a novel about cricket (板球).

**The Missing CD (Study 2)**

My sister Carly was standing in the hallway (门厅) at school. She was looking through her purse (小提包) for something. She kept asking herself where it was. I had no idea what she was looking for. She told me it was her CD. Without the CD, she couldn’t practice for the school play she was in. If she couldn’t practice, she would be kicked out of the play. I asked my sister where she might have left her CD. She was sure it was the gym (健身房). But when I suggested going there, she said she didn’t
want me to go with her. I asked her whether she really wanted to find her CD. She groaned (抱怨) and said I could come along.

When we got to the gym, we saw a blonde (金发女郎) girl. I asked who the girl was. The girl’s name was Shannon and she was unzipping (拉开) her purse. It was identical (同样的) to my sister’s. My sister immediately demanded to know why Shannon had taken her purse. Shannon said she did not understand what my sister was talking about. Then she accidentally (无意中) dropped her purse and a CD fell out! Shannon looked surprised and said she did not know how the CD had got in there.

I asked Shannon to give me her purse. She asked me why I wanted it. I explained I wanted to check something. I looked inside and saw “Carly” written on a name tag (名字标签). Then I looked inside the purse Carly was holding. My sister looked angrily at me and asked what I was doing. I looked inside Carly’s purse and saw “Shannon” on the name tag. Somehow the two girls had switched (交换) bags.

Shannon wanted to know where they could have switched the bags. It turned out that both Shannon and Carly had been having a meal in the school cafeteria (自助餐厅) and had put their bags on the same table. The two girls asked what they could do to thank me for sorting out (解决) their problem.

Directions: Now please report your name and group number to the voice recorder and then reproduce what you have read about the passage. You don’t have to use the original words but you should include all the main events in the passage. Please finish your retelling within four minutes.
Appendix F  Untimed Grammaticality Judgment
Pre/Post/Delayed Post-test

Name: ________  Group: ________

Instructions: In this section, you are expected first to decide whether each sentence is grammatically correct or incorrect for written English. Then, indicate your certainty of judgment by marking a level in a 6-point scale. 0 means that you are totally uncertain and 5 means that you are 100% confident. After that, indicate whether you have judged it by “rule” or by “feel”. Finally, if the sentence is ungrammatical, please correct it if you can. There is no time control in this section, so you can take your time to double-check your answers before turning over to the next page for a new item.

Study 1

1. My father takes something to read when he goes to the doctor’s.
   Grammaticality: A. correct  B. incorrect
   Certainty of judgment: very uncertain 0 1 2 3 4 5 very certain
   Judgment by: A. rule  B. feel
   If ungrammatical, correct it here: ____________________.

2. Everybody like Jack because he is good at telling stories.
   Grammaticality: A. correct  B. incorrect
   Certainty of judgment: very uncertain 0 1 2 3 4 5 very certain
   Judgment by: A. rule  B. feel
   If ungrammatical, correct it here: ____________________.

3. She remembers exactly everything that has happened to her.
   Grammaticality: A. correct  B. incorrect
   Certainty of judgment: very uncertain 0 1 2 3 4 5 very certain
   Judgment by: A. rule  B. feel
   If ungrammatical, correct it here: ____________________.

4. We’re going to the bookstore in John’s car this afternoon.
   Grammaticality: A. correct  B. incorrect
   Certainty of judgment: very uncertain 0 1 2 3 4 5 very certain
   Judgment by: A. rule  B. feel
   If ungrammatical, correct it here: ____________________.

5. The driver often remind passengers to take their belongings when they leave.
   Grammaticality: A. correct  B. incorrect
   Certainty of judgment: very uncertain 0 1 2 3 4 5 very certain
   Judgment by: A. rule  B. feel
   If ungrammatical, correct it here: ____________________.

6. The seaside here attracts a lot of tourists every summer.
   Grammaticality: A. correct  B. incorrect
   Certainty of judgment: very uncertain 0 1 2 3 4 5 very certain
   Judgment by: A. rule  B. feel
   If ungrammatical, correct it here: ____________________.

7. Every noon the sunlight comes in and lights up the whole room.
   Grammaticality: A. correct  B. incorrect
   Certainty of judgment: very uncertain 0 1 2 3 4 5 very certain
   Judgment by: A. rule  B. feel
   If ungrammatical, correct it here: ____________________.

231
8. My brother is fortunate to find a job that he love very much.
   Grammaticality: A. correct   B. incorrect
   Certainty of judgment: very uncertain  0 1 2 3 4 5 very certain
   Judgment by: A. rule     B. feel
   If ungrammatical, correct it here:__________________.

9. Only you can persuade Alice to give up her foolish idea.
   Grammaticality: A. correct   B. incorrect
   Certainty of judgment: very uncertain  0 1 2 3 4 5 very certain
   Judgment by: A. rule     B. feel
   If ungrammatical, correct it here:__________________.

10. We learn language by making and correcting mistakes.
    Grammaticality: A. correct   B. incorrect
    Certainty of judgment: very uncertain  0 1 2 3 4 5 very certain
    Judgment by: A. rule     B. feel
    If ungrammatical, correct it here:__________________.

11. Although I’m 20 years old, my grandpa still treat me like a child.
    Grammaticality: A. correct   B. incorrect
    Certainty of judgment: very uncertain  0 1 2 3 4 5 very certain
    Judgment by: A. rule     B. feel
    If ungrammatical, correct it here:__________________.

12. My wife used to ask, “Do you like my new shoes?”
    Grammaticality: A. correct   B. incorrect
    Certainty of judgment: very uncertain  0 1 2 3 4 5 very certain
    Judgment by: A. rule     B. feel
    If ungrammatical, correct it here:__________________.

13. Life in the future will be much easier.
    Grammaticality: A. correct   B. incorrect
    Certainty of judgment: very uncertain  0 1 2 3 4 5 very certain
    Judgment by: A. rule     B. feel
    If ungrammatical, correct it here:__________________.

14. The house belongs to my aunt but she seldom live here.
    Grammaticality: A. correct   B. incorrect
    Certainty of judgment: very uncertain  0 1 2 3 4 5 very certain
    Judgment by: A. rule     B. feel
    If ungrammatical, correct it here:__________________.

15. This is not a big shop but it offers more personal service.
    Grammaticality: A. correct   B. incorrect
    Certainty of judgment: very uncertain  0 1 2 3 4 5 very certain
    Judgment by: A. rule     B. feel
    If ungrammatical, correct it here:__________________.

16. We will have to wait all day unless the doctor work faster.
    Grammaticality: A. correct   B. incorrect
    Certainty of judgment: very uncertain  0 1 2 3 4 5 very certain
    Judgment by: A. rule     B. feel
    If ungrammatical, correct it here:__________________.

17. The lawyer always wears a suit in spring.
    Grammaticality: A. correct   B. incorrect
    Certainty of judgment: very uncertain  0 1 2 3 4 5 very certain
    Judgment by: A. rule     B. feel
    If ungrammatical, correct it here:__________________.
18. My parents want kicking me out of the house.
   Grammaticality: A. correct   B. incorrect
   Certainty of judgment: very uncertain  0  1  2  3  4  5  very certain
   Judgment by:     A. rule     B. feel
   If ungrammatical, correct it here:______________________.

Study 2

1. Lisa asked me yesterday if she could use my car.
   Grammaticality: A. correct   B. incorrect
   Certainty of judgment: very uncertain  0  1  2  3  4  5  very certain
   Judgment by:     A. rule     B. feel
   If ungrammatical, correct it here:______________________.

2. The scientists couldn’t figure out why had so many birds died.
   Grammaticality: A. correct   B. incorrect
   Certainty of judgment: very uncertain  0  1  2  3  4  5  very certain
   Judgment by:     A. rule     B. feel
   If ungrammatical, correct it here:______________________.

3. Wars never settle anything; they only lead to violence.
   Grammaticality: A. correct   B. incorrect
   Certainty of judgment: very uncertain  0  1  2  3  4  5  very certain
   Judgment by:     A. rule     B. feel
   If ungrammatical, correct it here:______________________.

4. Excuse me, can you please tell me where the nearest bank is?
   Grammaticality: A. correct   B. incorrect
   Certainty of judgment: very uncertain  0  1  2  3  4  5  very certain
   Judgment by:     A. rule     B. feel
   If ungrammatical, correct it here:______________________.

5. I want to ask the players why did they fail to win the game.
   Grammaticality: A. correct   B. incorrect
   Certainty of judgment: very uncertain  0  1  2  3  4  5  very certain
   Judgment by:     A. rule     B. feel
   If ungrammatical, correct it here:______________________.

6. Mary finally accepted Bruce as her life-long companion.
   Grammaticality: A. correct   B. incorrect
   Certainty of judgment: very uncertain  0  1  2  3  4  5  very certain
   Judgment by:     A. rule     B. feel
   If ungrammatical, correct it here:______________________.

7. Little Johnny feels the bag, curious to know what does it contain.
   Grammaticality: A. correct   B. incorrect
   Certainty of judgment: very uncertain  0  1  2  3  4  5  very certain
   Judgment by:     A. rule     B. feel
   If ungrammatical, correct it here:______________________.

8. You told Julia about the result, don’t you?
   Grammaticality: A. correct   B. incorrect
   Certainty of judgment: very uncertain  0  1  2  3  4  5  very certain
   Judgment by:     A. rule     B. feel
   If ungrammatical, correct it here:______________________.

9. I don’t mind where we go as long as there is sun, sea and sand.
   Grammaticality: A. correct   B. incorrect
10. I asked Mary could she tell me how to fill out the form.
   Grammaticality: A. correct   B. incorrect
   Certainty of judgment: very uncertain  0  1  2  3  4  5  very certain
   Judgment by:     A. rule        B. feel
   If ungrammatical, correct it here:_____________________.

11. We often provide our children with toys, footballs or basketballs to play with.
   Grammaticality: A. correct   B. incorrect
   Certainty of judgment: very uncertain  0  1  2  3  4  5  very certain
   Judgment by:     A. rule        B. feel
   If ungrammatical, correct it here:_____________________.

12. She asked the doctor why her hand was always shaking.
   Grammaticality: A. correct   B. incorrect
   Certainty of judgment: very uncertain  0  1  2  3  4  5  very certain
   Judgment by:     A. rule        B. feel
   If ungrammatical, correct it here:_____________________.

13. I’m afraid we will never know what is a UFO.
   Grammaticality: A. correct   B. incorrect
   Certainty of judgment: very uncertain  0  1  2  3  4  5  very certain
   Judgment by:     A. rule        B. feel
   If ungrammatical, correct it here:_____________________.

14. My grandparents have been married since 40 years.
   Grammaticality: A. correct   B. incorrect
   Certainty of judgment: very uncertain  0  1  2  3  4  5  very certain
   Judgment by:     A. rule        B. feel
   If ungrammatical, correct it here:_____________________.

15. I doubt whether anyone in my class has a higher IQ than Jane.
   Grammaticality: A. correct   B. incorrect
   Certainty of judgment: very uncertain  0  1  2  3  4  5  very certain
   Judgment by:     A. rule        B. feel
   If ungrammatical, correct it here:_____________________.

16. The English play what my students performed was a great success.
   Grammaticality: A. correct   B. incorrect
   Certainty of judgment: very uncertain  0  1  2  3  4  5  very certain
   Judgment by:     A. rule        B. feel
   If ungrammatical, correct it here:_____________________.

17. We are not sure whether will Tom come soon.
   Grammaticality: A. correct   B. incorrect
   Certainty of judgment: very uncertain  0  1  2  3  4  5  very certain
   Judgment by:     A. rule        B. feel
   If ungrammatical, correct it here:_____________________.

18. Susan didn’t want her parents to know what she was doing.
   Grammaticality: A. correct   B. incorrect
   Certainty of judgment: very uncertain  0  1  2  3  4  5  very certain
   Judgment by:     A. rule        B. feel
   If ungrammatical, correct it here:_____________________.

234
Appendix G  Instructions for Nonword Span Test

“请仔细听录音。当听到提示音‘Okay’的时候，无论您是否能够听懂，请立即将‘Okay’前边的词汇串重复出来。不需要考虑词汇的先后顺序，尽可能把所有的词汇都重复出来。如果某个词汇串没有跟上，请接着做下一个，直至老师宣布考试结束。”

“现在，请您先做个练习。”

“练习到此结束，现在开始考试。请向录音设施报出您的姓名和组别。您的姓名是...，您所在的组别为...”。”
Appendix H  Instructions for Digit Span Test

“首先，请向录音设施报出您的姓名和组别。您的姓名是…，您所在的组别为…。然后，请仔细听录音。当您听到提示音‘Okay’时，请立即将‘Okay’前边的数字串重复出来。不需要考虑数字间的先后顺序，尽可能把所有的数字都重复出来。如果某个数字串没有跟上，请接着做下一个，直至老师宣布考试结束。现在，考试正式开始。”
Appendix I  Listening Span Test

“在听力广度测试中，大家将听到成组的汉语句子。每组包括 3 到 6 个句子不等。在听到每个句子时，您要完成两项任务。一是判断该句子是否能够明确表达意思，讲得通。如果听起来符合逻辑，可接受，请按下“√”键；如果词不达意，听起来有问题，则按下“×”键。二是同时记住句子末尾的词汇。整组句子全部做完后，电脑屏幕上就会出现“回忆”一词。看到该词后，请将所听到的每个句子的尾词全部写到答题纸上。书写时，句尾词不分前后顺序。但需要注意的是，判断每个句子是否符合逻辑时的反应时间、判断的准确度、以及句尾词回忆的准确度都是评分的依据，三者同等重要。”

“现在，我们先做做练习。”

| 练习一 | 他应该参考别人的意见，谨慎作出决定。  
|        | *今天忙了整整一天，我累得说起来没完。 |
| 练习二 | *他从小就注重锻炼，所以就弄坏了身体。  
|        | 李先生工作很辛苦，但收入却不太理想。  
|        | *过去，北京的公交车站牌将要加注英文。 |
| 练习三 | 老人一想到过去的苦日子，就心酸难过。  
|        | 这些茶壶除了实用外，也很有艺术价值。  
|        | *除课本内，还可以借无关书籍回来参考。 |

“练习到此结束，现在开始考试。”

| 测试一 | 我们待人应以心胸宽广，对己要严格要求。  
|        | *经过漫长的急救，小狗总算救活了兽医。  
|        | *我国天然货物很多，种类只有煤和石油。 |
| 测试二 | *人人都买手机，因为它用起来很方便。  
|        | 这个节目非常有趣，适合全家老少观赏。  
|        | *我昨天工作到很晚，直到回家才十二点。  
|        | *这么重要的事给他做，我真的不能放心。  
|        | *大地震过后，各国都主动地捐款来购买。 |
| 测试三 | *他不仅夸奖顶头上司，还给上司加薪水。  
|        | 写论文引用资料时，应注明文献的出处。  
|        | *我出门没带钱，不幸遇到老友才没丢脸。  
|        | 美国东北部出现罕见的暴雪，损失惨重。  
|        | 老先生医术精湛，给患者留下深刻印象。  
|        | 如果不经常吸收新知识，很容易被淘汰。 |
| 测试四 | *我们必须认真修正五香十色的财经法规。  
|        | *这项工程已经竣工，可是无法准时完成。  
|        | 受国际油价飙升影响，物价呈上涨趋势。  
|        | 当义工不仅让自己充实，也使别人受惠。 |
| 测试五 | *老张疑心太重，不论人家说什么他都信。  
|        | *这次我考得不理想，因为时间不够充足。  
|        | 外面风声很紧，这件东西根本脱不了手。 |
| 测试六 | *公路暂时封闭，导致飞机不能正常起降。  
|        | *他每天一下课就从电影院跑出来看电影。  
|        | 这位候选人的政治见解很好，深得民心。  
|        | *我每次主动打扫卫生，妈妈都会训斥我。 |
他显然从错误中吸取了教训，又犯了错。
* 这列火车会到过那站，但是不会停下来。

** 测试七 **
爸爸常对我说，一分耕耘就有一分收获。
* 每到清明，总有五花八门的人采购礼物。
由于生活压力太大，他的睡眠质量不好。
他坚持每天读书，不断吸收并积累知识。

** 测试八 **
专柜售货员说这双鞋看起来非常适合你。
恐怖事件发生后，各国都加强防范措施。
* 看到热腾腾的鸡汤，无法令人五颜六色。
你千万不要错过这个千载难逢的好机会。

** 测试九 **
最近天气多变，你们出门要加件外套。
* 他学习用功，顺利地通过了这次考试。
网络计算机，应该采取有效防范措施。

** 测试十 **
只要有一点瑕疵，商品都可以退换顾客。
考试快到了，他正为考试做最后的冲刺。
* 虽然他有过失，你也犯不着当众表扬他。
在高考失败之后，他就变得失魂落魄了。

** 测试十一 **
只有天宇气里，想找到好工作非常不容易。
* 姐姐打电话，不论一高兴，就忘了时间。
为了爸爸身体健康，全家人都劝他戒酒。
因为事先没有规划，今天才会如此失败。

** 测试十二 **
经济不景气，想找到好工作非常不容易。
* 虽然他有过失，你也犯不着当众表扬他。
在高考失败之后，他就变得失魂落魄了。

** 测试十三 **
这个事情与我无关，对不起，都是我的错。
* 学校一定会对大家优异的表现得到批评。
三餐饮食要均衡，毕竟药补不如食补。
* 这些贝壳各有特色，都不值得慢慢欣赏。

** 测试十四 **
网络犯罪猖獗，应该采取有效防范措施。
考试快到了，他正为考试做最后的冲刺。
* 虽然他有过失，你也犯不着当众表扬他。
在高考失败之后，他就变得失魂落魄了。

** 测试十五 **
* 这列火车会到过那站，但是不会停下来。

** 测试十六 **
只要有一点瑕疵，商品都可以退换顾客。
考试快到了，他正为考试做最后的冲刺。
* 虽然他有过失，你也犯不着当众表扬他。
在高考失败之后，他就变得失魂落魄了。
### Appendix J  Answer Sheet for Listening Span Test

<table>
<thead>
<tr>
<th>姓名:</th>
<th>组别:</th>
</tr>
</thead>
</table>

#### 练习题

<table>
<thead>
<tr>
<th>练习一</th>
</tr>
</thead>
<tbody>
<tr>
<td>练习二</td>
</tr>
<tr>
<td>练习三</td>
</tr>
</tbody>
</table>

#### 测试题

<table>
<thead>
<tr>
<th>测试一</th>
</tr>
</thead>
<tbody>
<tr>
<td>测试二</td>
</tr>
<tr>
<td>测试三</td>
</tr>
<tr>
<td>测试四</td>
</tr>
<tr>
<td>测试五</td>
</tr>
<tr>
<td>测试六</td>
</tr>
<tr>
<td>测试七</td>
</tr>
<tr>
<td>测试八</td>
</tr>
<tr>
<td>测试九</td>
</tr>
<tr>
<td>测试十</td>
</tr>
<tr>
<td>测试十一</td>
</tr>
<tr>
<td>测试十二</td>
</tr>
<tr>
<td>测试十三</td>
</tr>
<tr>
<td>测试十四</td>
</tr>
<tr>
<td>测试十五</td>
</tr>
<tr>
<td>测试十六</td>
</tr>
</tbody>
</table>
Appendix K  Exit Questionnaire

结束问卷

非常感谢您在过去3个月中参加了本研究！为了更加全面地进行总结，我们请您认真填写下面这份问卷。您所提供的信息仅用于学术研究，绝对不会提供给学校校方或者其它任何第三方。请您坦诚回答每个问题。谢谢！

问卷中共有6个问题。只有回答完毕并提交了一个问题的答案，才能看到下一个题目。一旦提交，您将不能返回到上一题。因此，请您检查无误后再点击提交按钮。

1. 请首先写上你的姓名和所在的组别，然后点击提交。

2. 参与本研究的过程中，你是如何理解这些课的主要教学目标的？

3. 老师在课上都做了哪些具体工作？

4. 课堂活动过程中，你有没有注意到老师在纠正你或者其他同学的语言错误？如果当时注意到了，你能回忆起老师改正的语言错误都有那些吗？如果没有发现，请直接进入第5题。

5. 你认为老师纠错有用吗？为什么？

6. 对于这个研究项目，你还有其它的建议或者评论吗？
Appendix L  Coding Scheme for EI Data

1. An obligatory context is a place in a sentence where learners have to supply an exemplar of the target structure to make it grammatically correct. The obligatory contexts created by the elicitor and the obligatory contexts created by the participants while imitating distracter items were both counted.

2. Non-obligatory use refers to oversupply of an exemplar of the target structure in a context where the exemplar is ungrammatical. Eg. Last year, the earthquake shocks the whole world. Both the non-obligatory use in target sentences and distracter sentences were counted.

3. Third person singular subject includes “He”, “She”, “It” and such words as “Every/A students”, “Every/A children”, and “Its” at the very beginning of a sentence, due to the participants’ mispronunciation. Eg. Its tells.

4. Third person singular -s refers to the form of regular verbs in the simple present tense when the subject is 3rd person singular. Auxiliary verbs, such as “does”, “has”, “is” were excluded.

5. Attempts to produce an exemplar of the target structure were regarded as one obligatory context. A correct form achieved by multiple attempts was not coded as “correct suppliance” as the test was supposed to measure implicit knowledge.

6. Correct use of another verb after failing to produce the grammatical form of a verb in an obligatory context was not taken as “correct suppliance”.


Appendix M  Coding Scheme for OP Data

1. Only obligatory contexts created by the participants were considered in analyzing OP data.
2. Due to mispronunciation, “The presidents [referring to President Obama]”, “He’s/His” and “Its” in the position of subject were also regarded as 3rd person singular.
3. Any context where both direct and indirect questions applied was assumed to be an obligatory context for embedded questions.
4. Expressions like “what to do” were treated as a “non-obligatory context” for embedded questions.
5. Auxiliary verbs, such as “does”, “has”, “is” were not included in data analysis.
6. Attempts to produce an exemplar of the target structure were regarded as one obligatory context. A correct form achieved by multiple attempts was not coded as “correct suppliance” as the test was supposed to measure implicit knowledge.
7. Correct use of another verb after failing to produce the grammatical form of a verb in an obligatory context was not taken as “correct suppliance”.
8. Occasional supply of “-ed” form for a verb in describing President Obama’s daily routines was considered as an “obligatory context” without “correct suppliance”. Eg. At 6:30 every evening, he finished his work.
Appendix N  Scoring Scheme for UGJ Data

1. Only responses to target sentences were analyzed.
2. Wrong judgment of a grammatical sentence without indicating the error was scored “0”.
3. Wrong judgment of the target structure in a grammatical sentence was scored “0”.
4. “Wrong” judgment of the non-target structure in a grammatical sentence gained “1”.
5. Correct judgment of a grammatical sentence gained “1”.
6. Wrong judgment of an ungrammatical sentence was scored “0”.
7. “Correct” judgment of an ungrammatical sentence but wrong location of the error gained “0”.
8. Correct judgment of an ungrammatical sentence without indicating the error gained “0.5”.
9. Correct judgment of an ungrammatical sentence and correct location of the error gained “1”.
10. Correct judgment of an ungrammatical sentence, correct location of the error, and supply of the correct form gained “1.5”.
Appendix O  Results for the WM-Repair Relationship in the Corrective Recast Condition

**Table 1** The Corrective Recast Group’s WM and Repair Means for 3rd Person -s (n = 21)

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nonword span</td>
<td>3.52</td>
<td>.55</td>
</tr>
<tr>
<td>Listening span</td>
<td>.09</td>
<td>.52</td>
</tr>
<tr>
<td>Repair rate</td>
<td>.73</td>
<td>.23</td>
</tr>
</tbody>
</table>

**Table 2** The Corrective Recast group’s WM and Repair Correlations for 3rd Person -s (n = 21)

<table>
<thead>
<tr>
<th></th>
<th>Repair rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nonword span</td>
<td>.03</td>
</tr>
<tr>
<td>Listening span</td>
<td>-.11</td>
</tr>
</tbody>
</table>

*p < .05, **p < .01.

**Table 3** The CR Repairer and Non-Repairer Subgroups’ WM Means for Embedded Questions

<table>
<thead>
<tr>
<th></th>
<th>Nonword</th>
<th>Listening span</th>
</tr>
</thead>
<tbody>
<tr>
<td>CR Repairer Subgroup (n = 18)</td>
<td>3.37</td>
<td>.60</td>
</tr>
<tr>
<td>CR Non-repairer Subgroup (n = 6)</td>
<td>3.18</td>
<td>.42</td>
</tr>
</tbody>
</table>

Note. CR = corrective recasts.

**Table 4** The CR Repairer and Non-Repairer Subgroups’ WM Differences for Embedded Questions (n = 24)

<table>
<thead>
<tr>
<th></th>
<th>t</th>
<th>df</th>
<th>sig</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nonword span</td>
<td>.69</td>
<td>22</td>
<td>.49</td>
</tr>
<tr>
<td>Listening span</td>
<td>.79</td>
<td>22</td>
<td>.44</td>
</tr>
</tbody>
</table>

Note. CR = corrective recasts. *p < .05, **p < .01.
Appendix P  Results for the WM-Repair Relationship in the Implicit Recast Condition

Table 1 The IR Repairer and Non-Repairer Subgroups’ WM Means for 3rd Person -s

<table>
<thead>
<tr>
<th></th>
<th>Nonword</th>
<th>Listening span</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>IR Repairer Subgroup</td>
<td>4.09</td>
<td>.77</td>
</tr>
<tr>
<td>(n = 7)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IR Non-repairer Subgroup</td>
<td>3.44</td>
<td>.65</td>
</tr>
<tr>
<td>(n = 19)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. IR = implicit recasts.

Table 2 The IR Repairer and Non-Repairer Subgroups’ WM Differences for 3rd Person -s (n = 26)

<table>
<thead>
<tr>
<th></th>
<th>t</th>
<th>df</th>
<th>sig</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nonword span</td>
<td>2.16</td>
<td>24</td>
<td>.04*</td>
</tr>
<tr>
<td>Listening span</td>
<td>-.13</td>
<td>24</td>
<td>.90</td>
</tr>
</tbody>
</table>

Note. IR = implicit recasts. *p < .05, **p < .01.
References


