2016

Bridging the gap between declarative knowledge and procedural knowledge through metalinguistic corrective feedback

Wang, Qin

http://hdl.handle.net/2144/14568

Boston University
BRIDGING THE GAP BETWEEN DECLARATIVE KNOWLEDGE AND PROCEDURAL KNOWLEDGE THROUGH METALINGUISTIC CORRECTIVE FEEDBACK

by

QIN WANG

B.A., Liaoning Normal University, 1995
M.Ed., Boston University, 2002

Submitted in partial fulfillment of the requirements for the degree of

Doctor of Education

2016
Approved by

First Reader
Marnie Reed, Ed.D.
Associate Professor of Language Education

Second Reader
Bruce Fraser, Ph.D.
Professor of Linguistics and Education

Third Reader
Yingli Yang, Ph.D.
Associate Professor of English
University of International Business and Economics, School of International Studies
DEDICATION

To my husband Jianping Tang,

my child Chudi Tang (Mimi),

my parents Wenzhong Wang and Peifen Diao
ACKNOWLEDGMENTS

While this dissertation bears my name, truly this is collaborative work of many people around me during the years of my doctoral study. They have contributed to this work in different ways. Here I would like to formally thank them.

First and foremost, I am indebted to my dissertation committee—Dr. Marnie Reed, Dr. Bruce Fraser, and Dr. Yingli Yang. Dr. Reed has been a supportive and inspiring advisor and mentor throughout my graduate study at Boston University. Her intellect and dedication in language teaching has greatly motivated my research interest from my Master’s days. All the insightful ideas and details on revision I received from Marnie have greatly improved this work. I owe deep gratitude to Dr. Fraser for his generous consent to serve on my dissertation committee at an interim stage in its development. As an experienced and talented researcher, Dr. Fraser has provided me invaluable suggestions and constant encouragement in the research and writing process. My heartfelt thanks also extend to Dr. Yang. Her own research in corrective feedback has enlightened my research. Her constructive suggestions for revision have been highly beneficial for the refined version of this dissertation. Finally, I am thankful to Dr. Jeanne Paratore for her role in the initial stage of this dissertation. I have benefited very much from the guidance she provided in her research seminar.

My most sincere thanks also go to many colleagues and friends. Dr. Yaochen Deng has provided me with statistical help and support. Dr. Xianghua Tu, Dr. Cami Condie, Yu Wu, Julia Foster-Claudio have been important friends at BU. They have supported me with their love and friendship. Special thanks go to Cami and Xianghua for
reading my proposal and sharing their thoughts.

I would like to thank my colleagues and my students in China as well—Dr. Junyue Chang, Dr. Fengguang Liu, Dr. Yongqing Zhao, and the Class of 2015 at DUFL. Without them, this research would not be completed. Their participation and cooperation have played a vital role in this work.

Last but not least, I am deeply grateful for my dear family. My great thanks are for my husband, Jianping Tang, for his loving sacrifice and patience. I am proud of our decision to grow the family from two of us to three in the process of my doctoral study. My dear daughter, Mimi, has been a sweet and adorable support throughout the years. Thank you, Mimi! You motivate me to extend my endeavors to be a good mom.

Furthermore, I owe a deep debt of appreciation to my beloved parents. I could never have reached this stage of my academic journey without their immense love and support.
BRIDGING THE GAP BETWEEN DECLARATIVE KNOWLEDGE AND PROCEDURAL KNOWLEDGE THROUGH METALINGUISTIC CORRECTIVE FEEDBACK

QIN WANG

Boston University School of Education, 2016

Major Professor: Marnie Reed, Ed.D., Associate Professor of Language Education

ABSTRACT

Research in corrective feedback (CF) has developed over thirty years; however, little is known about how metalinguistic corrective feedback (MCF) assists the development of procedural knowledge. This quasi-experimental study sought to evaluate and compare the roles of oral MCF in proceduralizing English as a foreign language (EFL) learners’ declarative knowledge of gender referents and morphosyntactical structure—third person singular -s.

Sixty college EFL learners participated in the study. They were randomly assigned to the experimental group and the control group. All learners participated in 14 communicative form-focused activities that elicited the use of gender referents and third person singular morpheme -s in regular classes over 14 weeks. The experimental group received MCF on errors in use of the target structures. The control group received no MCF in use of the target structures. Pretests, posttests, and long delayed posttests administered 20 weeks after the instructional treatment assessed the participants’ declarative knowledge and procedural knowledge of both target structures. The metalinguistic knowledge test was employed to measure declarative knowledge. The
elicited oral imitation test was adopted to measure procedural knowledge.

Repeated measures ANOVA results indicated that the MCF group significantly increased their procedural knowledge of both target structures. Cohen’s effect size $d$ revealed that MCF exerted a bigger effect on proceduralizing EFL learners’ knowledge on third person singular morpheme –s than on gender referents. These results also demonstrated a relationship between the efficacy of oral MCF and the target structures as well as the interactional effect of test formats tapping into declarative knowledge and procedural knowledge. These findings shed light on the learnability of difficult structures through MCF and the necessity to employ metalanguage in EFL classrooms.
## TABLE OF CONTENTS

DEDICATION ................................................................................................................... iv

ACKNOWLEDGMENTS ................................................................................................. v

ABSTRACT ..................................................................................................................... vii

TABLE OF CONTENTS ................................................................................................. ix

LIST OF TABLES ........................................................................................................... xiii

LIST OF FIGURES ......................................................................................................... xv

CHAPTER ONE INTRODUCTION .................................................................................. 1

1.1 Rationale ................................................................................................................ 1

1.2 Research Questions ............................................................................................... 5

1.3 Significance ............................................................................................................ 5

CHAPTER TWO LITERATURE REVIEW ..................................................................... 7

2.1 Target Structures .................................................................................................. 7

2.1.1 Gender Referent Errors ................................................................................... 7

2.1.2 Third Person Singular Inflectional Morpheme -s Errors ................................. 12

2.2 Declarative Knowledge and Procedural Knowledge ........................................ 18

2.2.1 Distinction Between Declarative Knowledge and Procedural Knowledge .... 18

2.2.2 Declarative Knowledge and Procedural Knowledge Studies ....................... 21

2.3 Metalinguistic Corrective Feedback .................................................................. 26

2.3.1 Theoretical Framework of Corrective Feedback .......................................... 27

2.3.2 The Role of Metalinguistic Corrective Feedback .......................................... 33
2.3.2.1 The Role of MCF in Comparison with Other CF Types.................. 34
2.3.2.2 Empirical and Experimental MCF Studies............................. 35
2.3.2.3 Meta-analysis Studies of MCF......................................... 40
2.4 Summary .................................................................................. 43

CHAPTER THREE METHODOLOGY ......................................................... 46
3.1 Research Context ..................................................................... 47
3.2 Research Participants ............................................................... 48
3.3 Research Procedures ............................................................... 50
3.4 Target Structures ..................................................................... 52
3.5 Instructional Intervention .......................................................... 53
3.6 Testing Instruments .................................................................. 58
  3.6.1 The Elicited Oral Imitation Test ............................................. 58
  3.6.2 The Metalinguistic Knowledge Test ........................................ 60
  3.6.3 Questionnaire ...................................................................... 61
3.7 Scoring Procedure and Inter-rater Reliability .............................. 64
3.8 Data Analysis ........................................................................... 66

CHAPTER FOUR RESULTS ................................................................. 68
4.1 Analysis of Classroom Instructional Audio Data ............................. 68
  4.1.1 Coding Categories and Principles ......................................... 69
  4.1.2 Results of Classroom Instructional Recording ......................... 71
4.2 Analysis of the Test Results ....................................................... 75
  4.2.1 Gender Referents ................................................................. 76
CHAPTER FIVE DISCUSSION AND CONCLUSIONS

5.1 Summary of Research Findings

5.2 Efficacy of Metalinguistic Corrective Feedback

5.2.1 Explicitness and Noticeability of MCF

5.2.2 Uptake and Modified Output

5.2.3 Linguistic Targets

5.3 Declarative Knowledge and Procedural Knowledge

5.4 Pedagogical Implications

5.5 Limitations

5.6 Suggestions for Future Research

5.7 Final Conclusions

APPENDICES

Appendix 1 Informed Consent Form

Appendix 2 Pretest--Elicited Oral Imitation Test

Appendix 3 Pretest--Metalinguistic Knowledge Test
Appendix 4 Posttest--Elicited Oral Imitation Test……………………………………127
Appendix 5 Posttest--Metalinguistic Knowledge Test……………………………………128
Appendix 6 Short Exit Questionnaire……………………………………………………130
Appendix 7 Delayed Posttest--Elicited Oral Imitation Test…………………………131
Appendix 8 Delayed Posttest--Metalinguistic Knowledge Test…………………………132
Appendix 9 Student Copy of Elicited Oral Imitation Test………………………………134
Appendix 10 Statistic Results Through SPSS22…………………………………………135

BIBLIOGRAPHY………………………………………………………………………………147

CURRICULUM VITAE……………………………………………………………………………159
LIST OF TABLES

Table 3.1 Thematic Foci of Each Unit for Class Intervention ........................................48
Table 3.2 Bio-Data of the Participating Students ...............................................................49
Table 3.3 Treatment Activities to Practice the Target Features .......................................55
Table 4.1 Overall Distributions of Errors, Feedback Moves, Repair on Gender
Referents ......................................................................................................................72
Table 4.2 Overall Distributions of Errors, Feedback moves, Repair on Third Person
Morpheme -s ............................................................................................................73
Table 4.3 Gender Referents: Group Means and Standard Deviation in the Elicited Oral
Imitation Test .............................................................................................................77
Table 4.4 Gender Referents: Repeated Measures ANOVA in the Elicited Oral Imitation
Test Across the Three Testing Periods .....................................................................79
Table 4.5 Gender Referents: Group Means and standard Deviation in the Metalinguistic
Knowledge Test ..........................................................................................................80
Table 4.6 Gender Referents: Repeated Measures ANOVA in the Metalinguistic
Knowledge Test Across the Three Testing Periods .................................................81
Table 4.7 Third Person Singualr Morpheme -s: Group Means and Standard Deviation in
the Elicited Oral Imitation Test ................................................................................83
Table 4.8 Third Person Singualr Morpheme -s: Repeated Measures ANOVA in the
Elicited Oral Imitation Test Across the Three Testing Periods ................................84
Table 4.9 Third Person Singualr Morpheme -s: Group Means and Standard Deviation in
the Metalinguistic Knowledge Test ..............................................................................85
Table 4.10 Third Person Singular Morpheme -s: Repeated Measures ANOVA in the Metalinguistic Knowledge Test Across the Three Testing Periods..................87

Table 4.11 Effects of MCF on Gender Referents and Third Person Singular Morpheme -s in Different Tests (in Cohen's d effect size).............................................. 89

Table 5.1 Summary of Group Means Over Time.................................................94
LIST OF FIGURES

Figure 3.1 Overall Research Design................................................................. 51

Figure 3.2 Testing Procedure............................................................................... 63

Figure 4.1 Gender Referents: Group Means of Percentage Scores in the Elicited Oral
Imitation
Test..................................................................................................................78

Figure 4.2 Gender Referents: Group Means of Percentage Scores in the Metalinguistic
Knowledge Test..................................................................................................80

Figure 4.3 Third Person Singular Morpheme –s: Group Means of Percentage Scores in
the Elicited Oral Imitation Test.......................................................................83

Figure 4.4 Third Person Singular Morpheme –s: Group Means of Percentage Scores in
the Metalinguistic Knowledge Test.................................................................86
CHAPTER ONE
INTRODUCTION

1.1 Rationale

Despite having taken English as a Foreign Language (EFL) courses at middle and high schools for over six years, Chinese college EFL major students still have various phonological, lexical, and grammatical errors. Among them, two errors stand out as common and persistent. First, the learner uses third person singular personal pronouns *he* or *she* (and the relevant case forms of possessive determiners *his/her* or accusative pronouns *him/her*) interchangeably regardless of the gender of the referent. Second, the learner omits or misuses third person singular inflectional morpheme –*s* when simple present tense is needed.

While instructional methods have evolved from grammar translation to communicative language teaching, learners may have declarative knowledge but do not demonstrate procedural knowledge of the two linguistic forms. Declarative knowledge, also known as explicit knowledge in some research, is “all the conscious facts the learner has about the language and the criterion for admission to this category is the ability to articulate those facts. These may include some grammar rules, some vocabulary items, pronunciation rules, and so on” (Bialystok, 1978, p.72). Procedural knowledge, also known as implicit knowledge, is “the intuitive information upon which the language learner operates in order to produce responses (comprehension or production) in the target language” (Bialystok, 1978, p.72). Research has indicated that students may be able to achieve high scores on written tests and may be able to articulate the
metalinguistic knowledge about the linguistic features correctly; however, they fail to communicate accurately and fluently in spontaneous production (Hu, 2003; Lardiere, 1998a; Long, 2003). Chinese learners of English at the college level have been observed to display persistent gender referent and morpho-syntax errors in English classrooms (Chang & Ma, 2006; Dong & Li, 2011; Tao, 2003). It is thus necessary for EFL teachers and researchers to look for feasible ways to fill the gap between learners’ declarative knowledge and procedural knowledge of two target features.

Research in meaning-based and communicative language classrooms has shown that natural input is insufficient for learners to develop high levels of grammatical accuracy and error-free production (Harley, 1989; Harley & Swain, 1978, 1984; Lightbown & Spada, 1990). Interactionally modified input that arises from the negotiation of meaning between the language teacher and language learners in communicative contexts is proposed to be important for language development (Ellis, 2002; Long, 1983; Lyster, 2007; Norris & Ortega, 2000; Spada, 1997; Swain, 2000). Focus-on-Form Instruction (Form-focused Instruction or F-on-F), which combines form with meaning, “overtly draws students’ attention to linguistic elements as they arise incidentally in lessons whose overriding focus is on meaning or communication” (Long, 1991, pp. 45–46). F-on-F is claimed to help learners to notice the gap between interlanguage forms and target forms (Doughty & Williams, 1998; Swain, 1985). Research has demonstrated that F-on-F can improve learners’ language development in immersion and meaning-based classrooms (Day & Shapson, 2001; Ellis, 2001; Lyster, 2004, 2007; Norris & Ortega, 2000; Spada, 1997; Swain, 2000).
Corrective feedback (CF) is a reactive type of form-focused instruction (Yang & Lyster, 2010). It is the “feedback that provides learners with evidence that something they have said or written is linguistically incorrect” (Sheen, 2011, p.2). Many studies have demonstrated a positive effect of CF on errors made by English language learners (ELLs) from various first language (L1) backgrounds (Ammar & Spada, 2006; Doughty & Varela, 1998; Ellis, 2007; Han, 2002; Lyster, 1998, 2004; Lyster, Saito & Sato, 2013; Sato & Lyster, 2012; Sheen, 2004, 2007; Yang & Lyster, 2010). Recent CF studies primarily focus on the effectiveness of CF types. Researchers have demonstrated that output-prompting CF types, such as metalinguistic corrective feedback are superior to input-providing CF such as recasts (Carroll and Swain, 1993; Ellis, Loewen, and Erlam, 2006; Sheen, 2004, 2007, 2010). Metalinguistic corrective feedback (MCF) refers to “comments, information, or questions related to the well-formedness of the student’s utterance, without explicitly providing the correct form” (Lyster & Ranta, 1997, p. 47). A recast refers to a partial or complete reformulation of the learner’s erroneous utterance with the original meaning intact in a natural flow of conversation (Lyster, 1998; Li, 2010; Sheen, 2011). Compared to recasts, MCF provides learners with the opportunity to notice and self-repair their errors. Repair refers to learners’ provision of correct forms following feedback. Lyster (2004) even argues that the ideal way to help the learner convert declarative knowledge into procedural knowledge is through prompting self-correction and modified output. Based on the above conceptions, it is hypothesized in this study that MCF is a feasible strategy in language classrooms which helps to develop learners’ procedural knowledge of gender referents and third person morpheme –s.
Although CF studies have been carried out for over 30 years and have brought about many significant research findings, there are three research gaps in CF studies. Sheen (2011) points out that past and most recent research has only measured the learners’ declarative knowledge, leaving procedural knowledge untapped. To the best of my knowledge, only very few CF studies (Ellis, Loewen, & Erlam, 2006; Li, 2014; Loewen & Nabei, 2007; Sato & Lyster, 2012; Yang & Lyster, 2010) have attempted to measure learners’ procedural knowledge. The second research gap is the need for more studies to investigate the long-term effects of CF on language acquisition (DeKeyser, 2003; Mackey & Goo, 2007; Sheen, 2011). The majority of past and recent studies only examine the role of CF over a short period of several days or several weeks. The effects of CF over an extended period of time remain obscure. The third gap is the lack of studies comparing the efficacy of MCF on different types of linguistic forms in one study. Although target features are independent variables that greatly influence the effects of CF on language learning, researchers have uniformly focused on comparison of the effects of different CF types on one linguistic feature in one study rather than in relation to the different target structures (Ellis, 2007). The few exceptions are Ellis’ (2007) study that investigates the differential effects of CF on the two English grammatical features past tense –ed and comparative –er, Yang and Lyster’s (2010) study that explores the differential effects of CF on English regular and irregular past tense forms, and Li’s (2014) recent study that examined the effects of CF on Chinese classifiers and perfective –le.
1.2 Research Questions

This study explores the effects of Metalinguistic Corrective Feedback (MCE) on learners’ declarative and procedural knowledge of both gender referent and third person singular inflectional morpheme –s. It is hypothesized that MCF will reduce students’ gender referent and third person singular inflectional morpheme –s errors, reflecting the contribution of MCF to students’ development of their procedural knowledge of the two linguistic targets. This study proposes to answer the following questions:

Q1. Does oral metalinguistic corrective feedback increase the ability of Chinese university English majors to accurately assign gender referents? If yes, does the effect remain over an extended period of time?

Q2. Does oral metalinguistic corrective feedback increase the ability of Chinese university English majors to accurately produce the third person singular morpheme –s? If yes, does the effect remain over an extended period of time?

Q3. Does oral metalinguistic corrective feedback have differential effects on the acquisition of gender referents and third person singular morpheme –s?

1.3 Significance

The positive role of MCF in reducing grammatical and local errors has been evidenced in several studies, but the efficacy of MCF in reducing meaning-based lexical errors has not been intensely studied. This study facilitates the understanding of whether MCF has differential effects in reducing different types of errors made by Chinese EFL learners. That is it investigates whether MCF exerts different effects in reducing the global and lexical error of gender referents as it does on the local grammatical error of
third person inflectional morpheme -s.

In addition, the study attempts to fill the research gaps existing in the current research. It not only compares the longer-term effects of MCF on two linguistic forms but also examines the development of both declarative and procedural knowledge.

Last but not least, the study may contribute to the understanding of how language teachers can help their students convert declarative knowledge to procedural knowledge in second language (L2) classrooms. It is hoped that the findings of the study can contribute to pedagogical techniques for classroom error correction.
CHAPTER TWO

LITERATURE REVIEW

This chapter begins with an analysis of the two target features in this study--gender referent errors and the third person singular morpheme –s errors. Next, the distinction between declarative knowledge and procedural knowledge in language acquisition is illustrated. Studies regarding the two knowledge types are reviewed. Following this, discussion of various theories that are conducive to the facilitative role of CF in SLA is presented. Finally, empirical, experimental, and meta-analysis studies of MCF conclude the chapter.

2.1 Target Structures

2.1.1 Gender Referent Errors

In this study, gender referents refer to third person personal pronouns he/she and their case forms his/her/him. Possessive determiners his/her are categorized under gender referents following Celce-Murcia and Larsen-Freeman’s (1999) explanation:

The possessive pronoun forms are also part of the personal reference system in English. They perform two syntactic functions: a possessive form can serve as a possessive determiner before a noun, or it can replace an entire possessive NP.

This is Sheila’s book. ➔ This is her book. (p.298)

According to Ferris and Roberts (2001), the gender referent error falls in the category of lexical error, which includes errors of word choice and word form.

The gender referent error is one of the most stigmatizing lexical errors among
Chinese EFL learners (Dong & Jia, 2011; Dong & Li, 2011). Such errors are commonplace even among intermediate and advanced Chinese EFL speakers. The following speech is from a Chinese associate professor’s Advanced Oral English class in Tao’s (2003) analytical study of Chinese EFL learners’ misuse of gender referents. It documents the professor’s confusing uses of English gender referents when she introduced her college friend’s success at the beginning of the class:

Today we’d like to have a discussion about the topic: Should people be promoted according to ability? But before we have the discussion, I’d like to tell you about two university classmates of mine. One is now head of the foreign languages department of a key university of China. I had a phone talk to him the other day for two hours about his success. But to my great surprise he said she didn’t feel successful at all. I asked him why. She told me that she has actually changed a lot during the past years on his way from an ordinary teacher to head of the department. (p.45)

In the eight cases in use of third person singular pronouns, four of them are erroneous. All the errors are the misuses of she in place of he. Tao (2003) suggests that advanced Chinese EFL speakers make fewer gender referent errors than elementary and intermediate Chinese EFL learners. Advanced learners were often aware of the errors they make, but lower-leveled learners were not. Tao posits that advanced learners tend to replace he with she, which is contrary to lower-leveled learners’ tendency to misuse he for she.

Li and Han (2010) used the microgenetic method to observe and record the
change in a nineteen-year-old Chinese college EFL learner’s use of gender referents in four weeks. Their case study aimed to answer two research questions: 1. Do the erroneous uses of gender referents show a pattern? 2. Does the research participant reduce gender referent errors after receiving intensive visualization practice over a short period? The researchers used story-retell as a pretest to elicit the student’s use of gender referents. After the pretest, they provided the student with visualization aids as a means of gender referent practice for two 45-minute sessions in each of two weeks at the first research stage. In the following two weeks of the second research stage, the participant was asked to mentally visualize a man character and a woman character and to describe them. During the treatment, the researchers also asked the student to practice on her own. The activities the participant engaged in included oral descriptions of a third person and writing journals about a third person. Li and Han conducted a posttest immediately after the intervention and a delayed posttest four weeks later. Both tests were similar to the pretest. Both test results provided evidence for the reduction of gender referent errors. The highest error percentage of the six case forms of gender referents (subject pronouns he/she, possessive pronouns his/her, and object pronouns him/her) was 66.7% in the pretest when feminine object pronoun her was needed. Although the third person feminine object pronoun her remained the highest error percentage in the posttest, it dropped to 9.7% in the delayed posttest. Interestingly, the error incidence of both masculine subject pronoun he and masculine possessive pronoun his reduced to 0. Their case study sheds light on the possibility that with proper treatment learners’ gender referent errors can significantly diminish.
Likewise, Everett’s (1996) research showed a higher error rate of feminine gender referents than masculine gender referents. Everett conducted his study of gender referent errors made by Chinese college EFL learners to examine language transfer. Two tests were employed to elicit the students’ use of gender referents. Test One was a narrative task in which the participants were asked to describe each picture they saw in one English sentence. Each participant completed from 48 to 120 sentences in Test One. There was only one human image in each picture. Forty-one out of forty-four participants (i.e. 93%) had gender errors of third person pronouns. The rates of gender errors for the participants who had made the errors ranged from 2 percent to 26 percent, with an average percentage of 8.4. Test Two was a story-telling task in which the participants were supposed to tell the story of a little girl. There was also only one human character in the task. Forty-nine participants out of fifty-one (i.e. 96%) had gender errors of third person pronouns. The participants’ gender referent error rates ranged from 2.5 percent to 69 percent, with an average percentage of 25. Everett also found that the average error rate at 25% was higher when a feminine pronoun was needed than the error rate of 19% when a masculine pronoun was needed. This difference was significant at a p-value of 0.03. Although the tests to elicit the use of gender referents in this study are relatively easy compared to the other studies, the high percentage of errors is equally striking.

The result from Everett’s study is consistent with the analysis of Dong and Jia (2011). They analyzed gender referent errors made by Chinese college EFL learners from the corpus of SECCL (Spoken English Corpus of Chinese Learners). Their analysis showed that intermediate-level college students had a higher incidence of replacing she
with *he* than the other way around. The error rate reached 16.2% in the case where a feminine pronoun was needed; however, the error rate dropped to 3.47% where a masculine pronoun was needed. They pointed out that learners had more errors of subject pronouns and possessive pronouns than object pronouns. It was acknowledged that the actual error rate should be much higher since the researchers excluded errors when an antecedent such as *my friend* was used and thus the gender was made unclear due to students’ inconsistent use of gender referents throughout their speech. Dong and Jia (2011) emphasized that the confusing uses of gender referents they excluded from their analysis were much more than the ones they counted in their analysis.

In order to further test the universality of this error pattern among their Chinese college English learners and to determine the factors in relation to error incidences, the researchers assigned Chinese-English oral translation tasks to their fifty-eight third-year college English major students. The participants were asked to listen to sixteen Chinese stories and to translate them into English. The tasks vary in difficulty in the aspects of word frequency, complexity of sentence structures, and the story length. Some stories had only one male or one female character. Some had two characters of both genders. A few had multiple characters of both genders. The researchers found that the participants had more errors in easy translation tasks than in difficult translation tasks. The students had the highest gender referent error rate of 24.64% in a short and easy translation task in which there was only one character. They had the lowest error rate of 10.13% in a long and difficult translation task in which there were two characters of both genders. This finding indicates that participants’ gender referent error rates have very little to do with
the difficulty of the tasks. The researchers suggest that improving Chinese EFL learners’ monitoring in speech production of gender referents will help them to reduce gender referent errors.

Altogether, these studies point to the seriousness of Chinese EFL learners’ erroneous use of gender referents. The studies reviewed here evidence the four characteristics of learners’ gender referent errors. First, intermediate and low-level Chinese EFL learners have more errors of feminine pronouns in compulsory contexts where female referents are involved than masculine pronouns. Second, advanced learners tend to replace he with she, but lower-level learners tend to replace she with he. Third, learners have more errors of subject pronouns and possessive pronouns than object pronouns. Fourth, learners’ erroneous gender referent use has very little to do with the difficulty of the tasks eliciting their gender referent use. In addition, the research findings suggest that learners’ erroneous use of gender referents can be treated.

\textit{2.1.2 Third Person Singular Inflectional Morpheme –s Errors}

Unlike the typical gender referent errors among Chinese speakers, third person singular inflectional morpheme –s error seems universal among many non-native English language learners irrespective of their L1 (first language) backgrounds and their L2 (second language) proficiency levels. Researchers identify the error as one of the most serious and persistent problems for native speakers of Russian (Ionin & Wexler, 2002), Turkish (Haznedar, 2001; White, 2003), Basque, Spanish (Garcia Mayor & Villarreal Olaizola, 2010), and of course Chinese (Lardiere, 1998a, b; Chang and Ma, 2006). In
In general, there is consensus among these researchers that grammatical morphemes of tense are difficult structures for L2 learners to acquire.

Garcia Mayor and Villarreal Olaizola (2010) investigated 78 Basque-Spanish bilinguals at the secondary school level developing morphology in English as a third language. There were four morphemes in this study: third person singular morpheme –s, past tense –ed, auxiliary be, and copular be. The participants were divided into two groups—the content and language integrated learning (CLIL) group and the mainstream (non-CLIL) group. CLIL students received regular English instruction for three hours a week in addition to four to five hours of content teaching, such as religion and geography in English. Non-CLIL students only received regular school-wide English teaching for three hours a week. Both groups were asked to narrate a picture story and their narrations were audio-recorded twice over three years. The researchers found that the participants made the most errors in third person –s morpheme among all the four morphemes at both times. The percentages of the omitted third person morpheme were 71% at Time 1 recording and 22% at Time 2 recording for the CLIL group, 72% and 12% respectively for the two recordings for the non-CLIL group. Despite the extra English input the CLIL group received in content teaching, the researchers did not see significant group differences of the two recordings of third person –s morpheme.

Lardiere (1998a, b) examined the acquisition of L2 English inflectional morphemes by an adult Chinese speaker, Patty, in her longitudinal case study. Patty started to live in the USA at the age of 22 and got her bachelor’s and master’s degrees in American universities. The researcher audio recorded three spontaneous and naturalistic
conversations with Patty over eight years. At the age of thirty-two, ten years after Patty first settled down in the USA, she was first recorded. Eight-and-a-half years later when Patty was 41, the second and the third recordings were taped about two months apart. By then Patty had been married to a native-English speaker for about six years and had spent the intervening years since the first recording in a nearly exclusively English-speaking environment. She spoke entirely English at home and at work. The only chances she spoke Chinese were her occasional telephone conversations and meetings with her cousin in another state in the country. Patty estimated her proportion of usage of English (vs. Chinese) as ninety-five to ninety-eight percent (vs. two to five percent of Chinese use).

After transcribing and analyzing the data from the three recordings respectively about 34-minute long, 75-minute long, and 31-minute long, the researcher documented that the incidence of tense morphology in Patty’s spontaneous oral production was about 35% of obligatory contexts. In respect to third person singular morpheme (excluding the copula and auxiliary *be*), the correct suppliance rates in compulsory contexts were respectively 9.4%, 17%, and 12% at the three recordings. The study suggests that the increasing length, depth, and intensity of exposure to English environment at home and at work did not contribute to the development of third person inflectional morpheme –s. The researcher attributed Patty’s inaccurate use to the “fossilized steady state” of her English L2 morphology development.

As a part of their investigation of Chinese college EFL learners’ acquisition of third person singular inflectional morpheme –s and auxiliary verb *is*, Chang and Ma (2006) analyzed six intermediate and advanced Chinese EFL learners’ acquisition of third
person singular inflectional morpheme –s. Three of the participants were non-English major freshmen. The other three were a first-year non-English major graduate student, a first-year English major graduate student, and a first year doctoral student in linguistics who also had five-year English teaching experience. The researchers recorded their oral interviews of the participants for 30 to 40 minutes every two months from September 2004 to July 2005. In order to elicit the participants’ use of third person -s morpheme, the researchers interviewed them on the topic of a third person every time. They found that the four non-English major participants’ acquisition of third person singular morpheme –s was not satisfactory. Their accuracy percentage was under 21%. The doctoral student had the highest accuracy rate of 70.8%. One of the non-English major freshmen had the lowest accuracy rate of 14.3%. Apparently, the participants did not improve their accuracy level in third person singular morpheme –s in the absence of focused error treatment although they did get their regular formal English instruction over the ten-month research. To further investigate whether Chinese college EFL learners had the same errors in written English, they asked three groups of participants to write a 30-minute timed essay entitled my mother. The first group was 172 non-English major freshmen. The second group was 123 non-English major sophomores. The third group was 118 English major juniors. Compared with the accuracy rate from oral production, the participants had a high accuracy rate in writing. The analysis indicated that the English-major participants who were more competent English learners and had received more years of formal and intensive English instruction tended to have fewer errors than the others. The three groups had accuracy rates of using third person –s morpheme at
57.1%, 67.1%, and 80.5% respectively. The researchers attributed the higher accuracy rate to the ease of writing tasks compared to the speaking tasks. They posited that students faced more time-constraint pressure processing oral information than processing written information. It is also speculated that third person singular morpheme –s errors would remain if they were left unattended to.

Chen and Zhang (2008) investigated the roles of corrective feedback in the forms of recasts and elicitation in young Chinese learners’ acquisition of third person morpheme –s. Altogether 30 second-year middle school students were randomly assigned to recast, elicitation, and control groups. The research participants learned the target feature two months before the research was conducted, but they did not internalize the knowledge according to the researchers. The students took the pretest on the first day of the research and then received recasts and elicitation treatment in the following eight days, immediately followed by the posttest. The delayed posttest was conducted two weeks later. Picture-cued activities were used as both the testing instrument and treatment activities. The tests results showed that the recast group outperformed the elicitation group and the control group in both the immediate posttest and the delayed posttest. Although the performance of the recast group was better than the elicitation group in the immediate posttest and the delayed posttest, Pearson correlation coefficient indicated an insignificant correlation between the recast group’s uptake and their performances in the delayed posttest. The researchers believed that uptake of the recast group cannot indicate the acquisition of the 3rd person morpheme –s. In contrast, a strong correlation was found between the uptake of the elicitation group and their test performances. Chen and Zhang
claimed that prompted output following elicitation correction was an indication of the
development of third person morpheme –s in the elicitation group.

All the research results reviewed previously are in line with the findings from
Goldschneider and DeKeyser’s meta-analysis (2001). They pooled 12 studies, all of
which investigated ESL learners’ acquisition of 6 target structures: third person
morpheme –s, articles a, an, the, present progressive –ing, possessive ’s, regular past –ed,
and plural –s. Present progressive –ing received the highest accuracy percentage of 99.04.
Third person morpheme –s received the lowest accuracy percentage. The accuracy
percentages of third person morpheme –s range from 0.767 to 66.61 in the 12 studies.
The accuracy percentage of third person morpheme –s ranks the fifth on the accuracy
percentage list of the six structures in only 3 studies. It is the last on the accuracy
percentage list in the remaining 9 studies involved in the meta-analysis. It is manifest that
English learners have serious problems using third person morpheme –s correctly.

Overall, these studies indicate that third-person singular morpheme –s is a
common and persistent error among non-native English speakers at different levels and
from different first language backgrounds. Even highly proficient L2 speakers have
difficulty in accurately using this grammatical morpheme in communicative contexts.
They commit more errors in speaking than in writing. Research shows that they make
little progress over time of their formal English instruction or over their stay in an
English environment. However, F-on-F practice and modified output may be correlated
with learners’ acquisition of third person morpheme –s.
2.2 Declarative Knowledge and Procedural Knowledge

In the research of SLA, it is necessary to distinguish two types of linguistic knowledge—declarative knowledge (also known as explicit knowledge) and procedural knowledge (also known as implicit knowledge). Bowles (2011) points out “When linguists and L1 acquisition researchers refer to linguistic knowledge, they typically refer to linguistic competence (i.e., implicit/procedural knowledge). However, in many empirical SLA studies, learners’ knowledge of targeted structures in a language is often measured through tests that favor the use of explicit/declarative knowledge” (p.252). The distinction between the two types of knowledge thus contributes not only to our understanding of learners’ interlanguage development but also to the valid measure of their learning outcomes.

2.2.1 Distinction Between Declarative Knowledge and Procedural Knowledge

Bialystok (1981) explains procedural (implicit) knowledge and declarative (explicit) knowledge in this way:

implicit (procedural) knowledge is descriptive in that the structure of the knowledge is not apparent apart from its application; explicit (declarative) knowledge is explanatory in that its logical basis is understood independently of its application.  (p. 34)

According to Bialystok (1978), declarative knowledge is the learners’ knowledge of all the conscious facts about the language and the learners’ ability to articulate those facts. It is the knowledge of what. Procedural knowledge is the learners’ intuitive application of
the linguistic knowledge to produce responses in the target language. It is the knowledge of how. Declarative knowledge and procedural knowledge are processed in different ways, either automatic or controlled. Bialystok (1979) postulates that automatic rules may be represented in procedural knowledge while structural rules that are not under implicit control may reside in declarative knowledge.

Ellis (2009) identified four criteria to distinguish declarative and procedural knowledge. The four criteria are as follows:

1. degree of awareness;
2. the time available for producing a response;
3. the focus of attention (on meaning or on form);
4. the utility of metalanguage in producing a response. (p.338)

According to this study, the use of declarative knowledge requires “high awareness, ample response time, a focus on form and opportunity to use metalanguage” (p.338), whereas the use of procedural knowledge requires “low awareness, limited response time, a focus on meaning and little opportunity to use metalanguage” (p.338). It is suggested that the tests of procedural and declarative knowledge have different foci as to meaning vs. form and the procedural knowledge should be operated by “feel” in “real time” while the declarative knowledge should be the presentation or application of metalinguistic knowledge under no time constraint (Ellis, 2007). Bowles (2011) summarizes the features of declarative knowledge as explicit, “conscious”, “highly variable”, and “only accessible through controlled processing”; on the contrary, procedural knowledge is implicit, intuitive, “automatic”, “variable in a limited and systematic way”, and “available in
fluent, spontaneous language use” (p.251).

Bialystok (1979) holds that declarative knowledge and procedural knowledge are on a continuum; in other words, declarative knowledge can be developed to procedural knowledge. It is believed that procedural knowledge is developed through exposure to communicative language use and facilitated through functional practice. She posits that declarative knowledge arises when learners focus on the language code and engage in formal practice. Bialystok’s standpoint is consonant with that of Ellis (1993), who asserts that declarative knowledge can be converted into procedural knowledge through practice if the learners are developmentally ready to acquire the linguistic feature.

To achieve native-like proficiency, L2 learners go through three stages: 1. declarative knowledge stage, 2. procedural knowledge stage, and 3. automatizing stage (DeKeyser, 2007a,b). At the first stage, learners develop declarative knowledge of the language. However, they may not necessarily be able to apply the knowledge correctly. At the second and the third stages, learners are developing procedural knowledge, which enables them to subconsciously use the language correctly and eventually to automatize their knowledge and use of the language. DeKeyser (2003) states that declarative knowledge can become procedural knowledge in the sense that “learners can lose awareness of the structure over time, and learners can be aware of the structure of implicit knowledge when attempting to access it” (p.315). It is manifest that the development of procedural knowledge takes longer time than the development of declarative knowledge. Nevertheless, repeated practice generated from output-pushing CF types such as MCF can help learners to proceduralize declarative knowledge (Lyster, 2004).
2.2.2 Declarative Knowledge and Procedural Knowledge Studies

To a great extent, Ellis’ four criteria have been validated by research as operational to measure declarative and procedural knowledge (Hulstijn & Hulstijn, 1984; Han & Ellis, 1998; Ellis, 2005; Erlam, 2006; Elder & Ellis, 2009; Loewen, Erlam, & Ellis, 2009). An early study by Hulstijn and Hulstijn (1984) investigated the influence of time pressure and focus of attention on the correct use of two Dutch word order rules. 32 adult Dutch learners performed story-retelling tasks after listening to passages in Dutch as L2 under four conditions. In the Information/Fast condition, the participants had to pay attention to the information they were asked to reproduce later and at the same time speak as fast as they could. In the Information/Slow condition, they could take as long as they needed. In the Grammar/Fast condition and Grammar/Slow condition, they had to pay attention to grammatical accuracy. The researchers found that time pressure resulted in shorter responses than no time pressure and had more influence when the attention was on grammar than on information. Dutch learners in both Grammar conditions produced longer responses (which were measured in seconds) than those in the Information conditions despite the time pressure. Although no main time effect was found in this study, focusing on form took time and produced less information (which was measured by the scores of correct information). To be specific, the two focusing on grammar conditions took 30% longer time and produced 14% less information than the two information conditions. Another finding is that focus of attention on grammar resulted in the participants’ better performance of the two target forms, which indicated the increased monitoring in the participants’ speech production. Hulstijn and Hulstijn (1984)
concluded that time alone is not a necessary condition for self-correction, however, focus on form demands time for successful self-correction.

Hulstijn and Hulstijn continued their study to examine the relationship between declarative knowledge of grammar rules and the procedural knowledge of applying the rules. The same participants were asked to take a sentence correction test and after a couple of weeks were interviewed in Dutch about their ability to verbalize the grammar rules of the same two Dutch linguistic forms as in the first part of the research. They compared the results with those from the first experiment of the research. They found that there were no significant differences in performance scores between the participants with declarative knowledge and those without declarative knowledge, which indicated that declarative knowledge of grammar rules did not guarantee the ability to apply the rules. They posit that human beings have limited capacities for information processing. When EFL learners do not yet use L2 structural rules automatically, they need to pay extra attention to the rules in their speech production in addition to their attention to other phenomena such as informational aspects of their message. If they are under a time constraint, they are likely to run out of the capacity to simultaneously control the structural features of linguistic items. It is thus highly likely that learners commit structural errors under such circumstances.

Han and Ellis (1998) also made an attempt to explore the instruments of measuring L2 learners’ declarative and procedural knowledge of verb complementation structures. They proposed two criteria to differentiate the two types of knowledge:

Implicit knowledge is easily accessed in tasks that call for fluent language
performance. In contrast, explicit knowledge can be accessed only with controlled effort and, thus, is typically used in tasks that allow for careful planning and monitoring. Whereas implicit knowledge is unanalyzed, and constantly held without awareness, explicit knowledge is analyzed and model-based, and thus represents consciously held insights about language (p.6).

Four tests were administered to 48 adult ESL learners in a university intensive English language program. An oral production test and a timed grammaticality judgment test were designed to measure procedural knowledge. An untimed grammaticality judgment test and an interview were designed to measure declarative knowledge. A Principle Component Factor Analysis was performed and revealed the possibility of designing tests to measure declarative knowledge and procedural knowledge. Pearson product moment correlations were also calculated to address the role of declarative and procedural knowledge in general language proficiency. The results showed that declarative knowledge only played an insignificant role in general language proficiency.

Before refining the criteria to distinguish declarative knowledge and procedural knowledge into four, Ellis (2005) used 7 criteria to develop a battery of tests to investigate the relationship between declarative knowledge and procedural. He designed five tests to examine 17 English grammatical features, including third person morpheme –s, one of the target structures in this study. In addition to the four criteria reviewed earlier, he also used another three criteria, namely systematicity, certainty, and learnability. The five tests are imitation test, oral narrative test, timed grammar judgment test, untimed grammar judgment test, and metalinguistic knowledge test. Twenty native
speakers of English and ninety-one L2 English learners were involved in this study. Ellis found that time-pressured tests demand learners to rely on procedural knowledge, whereas tests without time constraints permit learners to draw on both types of knowledge. In his view, the imitation test and the metalinguistic knowledge test are respectively best to measure procedural knowledge and declarative knowledge.

Erlam (2006) further tested the validity of the elicited imitation test as a measure of procedural knowledge in her empirical study in the following year. She used a reconstructive elicited imitation test to measure 111 participants’ procedural knowledge. 20 were native speakers of English and the rest were ESL learners in New Zealand. Of the non-native English-speaking participants, 78% were native Chinese speakers. The test required the participants to first process statements for meaning and then to repeat those grammatically correct statements and to correct those ungrammatical statements. Before repeating the statements, the participants were asked to do the “Belief Questionnaire”. According to Erlam, this demanded a primary focus on meaning rather than on form. It also prevented participants from relying on memory and repeating what they heard verbatim as in other regular imitation tests. The reconstructive elicited imitation test required participants to process, rather than repeat verbatim, language stimuli. Third person inflectional morpheme –s was also examined as one of the seventeen target structures elicited by the reconstructive imitation test and the oral narrative test. The result showed that participants performed better on the spontaneous language production task than on the elicited imitation task. She claims that test takers’ ability to correct ungrammatical items is a better indication of procedural knowledge than the ability to
repeat correct items. The researcher concludes that the reconstructive elicited oral imitation test is a valid and reliable measure of L2 procedural knowledge.

Bowles (2011) recently duplicated Ellis’s (2005) study and provided support for the battery of tests employed in Ellis’ study as valid measures of declarative knowledge and procedural knowledge. 30 Spanish speakers and learners completed the five tests reported in Ellis’ study. 10 were native Spanish speakers who immigrated to the USA as adults from different countries in Latin America and South America. 10 were L1 English speakers and L2 Spanish learners from intermediate-level Spanish classes in an American university. 10 were heritage learners of Spanish who were raised in bilingual English-Spanish families.

Different One-Way ANOVAs were performed and statistically significant between-group differences were found on all five tests. L2 learners scored highest on the metalinguistic knowledge test (72.4%) and the untimed grammatical judgment test (66.9%), which were designed to measure declarative knowledge. They scored under 50% accuracy in the oral imitation test, the oral narrative test, and the timed grammar judgment test, which were designed to measure procedural knowledge. Contradictorily, heritage learners scored lowest on the metalinguistic knowledge test (57.4%) and highest on the oral narration test (95.9%). After comparing the scores on the timed versus untimed grammatical judgment tests, Bowles found that “time was a greater factor for L2 learners than for either HL learners or NSs of Spanish” (p. 262). She posited that L2 learners relied more heavily on declarative knowledge to complete the untimed grammatical judgment test than their counterparts did. Overall, the study provided
support for the battery of tests as reliable and valid measures of different types of knowledge.

To conclude, it is crucial to differentiate declarative knowledge from procedural knowledge in SLA research. Declarative knowledge and procedural knowledge are different in four aspects—degree of awareness, response time, focus of attention on form or on meaning, and use of metalinguistic knowledge in response (Ellis, 2009). Possession of declarative knowledge does not guarantee the development of procedural knowledge, let alone the conclusion that learners acquire the target structures. Although different, declarative knowledge can be developed into procedural knowledge through communicative practice and use of the target language (Bialystok, 1979; DeKeyser, 2003, 2007a, b; Ellis, 1993; Lyster, 2004). Furthermore, learning outcomes demand measurement in relation to knowledge types. The measures of both types of knowledge are indispensible for well-designed research that investigates learners’ acquisition of target forms. The metalinguistic knowledge test is found to be the best to test declarative knowledge and the elicited oral imitation test is found to be a valid measurement of procedural knowledge (Bowles, 2011; Ellis, 2005; Erlam, 2006).

2.3 Metalinguistic Corrective Feedback

Corrective feedback (CF) was first defined as “any reaction of the teacher which clearly transforms, disapprovingly refers to, or demands improvement of the learner utterance” (Chaudron, 1977). Sheen (2011) later defined CF as any feedback that provides learners with evidence that something they said or wrote is linguistically
incorrect. The efficacy of corrective feedback has been widely studied over the past decades ever since Chaudron’s study in 1977. Empirically, research has reached a consensus that MCF has a positive role in language development. Theoretically, cognitive and sociocultural theories have provided support for the beneficial role of CF in SLA. In this section, the theoretical framework of CF and studies in support of the beneficial role of MCF will be reviewed.

2.3.1 Theoretical Framework of Corrective Feedback

SLA theories have provided theoretical support for the positive role of CF from different aspects. Lyster, Saito and Sato (2013) assert that “theoretical perspectives that run the gamut from cognitively to socially oriented suggest that CF is not only beneficial but may also be necessary for moving learners forward in their L2 development” (p.9). I will discuss the theoretical framework of CF from the perspectives of cognitive theories and sociocultural theories in this section.

Interaction Hypothesis

Long developed the Interaction Hypothesis based on Krashen’s Input Hypothesis. In the Input Hypothesis, Krashen (1981) proposes that language learners’ exposure to sufficient and comprehensible target language input is “the only causative variable in SLA” (p.57). Of the two types of input, Krashen affirms the important role of positive evidence in SLA but ignores the facilitative role of negative evidence. Krashen’s argument is challenged in Long’s early Interaction Hypothesis. Long (1983) contends
that negotiation of meaning and form as well as linguistic modification in a meaningful and interactional context arise when learners have communication breakdowns and try to make their speech comprehensible. He attaches importance to negative evidence in SLA.

According to Gass (2003), positive evidence is the input that consists of well-formed sentences and speech samples, while negative evidence is the input that is provided to learners regarding the inaccuracy of an utterance. CF is a typical example of negative evidence provided to learners regarding the incorrect use of the target language in an interactional context. Long (1996) holds a positive view on the facilitative role of CF as a form of negative evidence in his updated Input Hypothesis. He discusses learners’ negotiation in communication as follows:

the process in which, in an effort to communicate, learners and competent speakers provide and interpret signals of their own and their interlocutor’s perceived comprehension, thus provoking adjustments to linguistic form, conversational structure, message content, or all three, until an acceptable level of understanding is achieved (p. 418).

Long’s position has been supported by SLA researchers such as Panova and Lyster (2002) and Spada and Lightbown (2008). On the basis of their analysis of several studies, Panova and Lyster (2002) lend their support in this way:

although a great deal of L2 learning takes place through exposure to comprehensible input, learners may require negative evidence (i.e., information about ungrammaticality), in the form of either feedback on error or explicit instruction, when they are not able to discover through exposure alone how their
interlanguage differs from the L2. (p.573)

**Output Hypothesis**

Based on the findings of empirical studies in content-based and language immersion contexts, Swain (1985, 1995) proposed that comprehensible input is not sufficient for language learning and output is a necessary complement to successful language acquisition. She found from studies of immersion and communicative classrooms (eg. Harley, 1989; Harley & Swain, 1978, 1984; Lighbown & Spada, 1990, 1994) that students did not achieve speech accuracy at a grammatical level despite their exposure to comprehensible input. Obviously, comprehensible input alone is inadequate for language development. She argued that lack of opportunities for pushed output led students to their failure to achieve grammatical accuracy.

Swain (1995) distinguished three functions of output: a noticing function, a hypothesis testing function, and a metalinguistic function. First, learners will become aware of the gap between their interlanguage and the target language in their output process such as speaking and writing. Second, output serves as opportunities for learners to test their hypotheses about the linguistic accuracy and comprehensibility and to modify their output in response to interlocutor’s feedback. Third, learners will reflect on their target language use in output processes and “their output serves a metalinguistic function, enabling them to control and internalize linguistic knowledge” (Swain, 1995, p.126).
**Noticing Hypothesis**

Schmidt (1990, 1994) proposed the Noticing Hypothesis which claimed noticing is a conscious process necessary for learning. Drawing on self-reported and tape-recorded data from a case study of Schmidt himself as a Portuguese learner in Brazil (Schmidt & Frota, 1986), he found that he failed to acquire the forms he did not consciously attend to in the input but he was able to produce the forms he noticed during his interaction with native speakers. Schmidt (1990) suggests that there is a close connection between noticing and emergence in production. Learners must notice the mismatch between what they can say and what they want to say. Under the framework of the Noticing Hypothesis, corrective feedback provides learners with opportunities to notice the gap between the target forms and their interlanguage forms and to modify their erroneous production, thereafter prompting self-repair and facilitating their language development.

**Skill Acquisition Theory**

Skill Acquisition Theory (Anderson, 1983, 2005; Johnson, 1996) postulates that SLA is similar to other cognitive skills acquisition. Language learners go through three stages to acquire L2. They first develop declarative knowledge, then proceduralize it, and finally automatize it (DeKeyser, 2007a, b). Repeated practice plays a crucial role in the gradual shift from declarative knowledge to procedural knowledge, i.e. proceduralization (DeKeyser, 1998, 2007a; Lyster & Sato, 2012). DeKeyser defined practice as “specific activities in the second language, engaged in systematically, deliberately, with the goal of developing knowledge of and skills in the second language” (2007a, p. 1). Here CF
comes into play. Despite input-providing CF types or output-prompting CF types, CF provides controlled practice opportunities for language learners to acquire language skills. In the aspect of converting declarative knowledge into procedural knowledge, Lyster (2004) argues that output-prompting CF types are more beneficial than input-providing CF types for the more opportunities of practice provided by the former than the latter.

Counterbalance Hypothesis

Lyster and Mori (2006) proposed the Counterbalance Hypothesis after their comparative analysis of French immersion classrooms and Japanese immersion classrooms. They argue that interlanguage development can be facilitated if pedagogical intervention acts as the counterbalance to the predominant classroom orientation. The 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 than instructional activities and interactional feedback that are congruent with the predominant communicative orientation (Lyster & Mori, 2006, p. 294).

In this view, CF plays a counterbalance role in a content-focused or a communicative classroom and is likely to be proven more effective than the pedagogical intervention that is in line with the prevailing classroom orientation.
Sociocultural Theory

Sociocultural Theory is built on the work of Vygotsky (1978, 1986). In the sociocultural view, higher-order mental activities such as language learning are socially mediated operations and an important mediation tool is social interaction. There are two key component concepts in Sociocultural Theory that are closely related to CF. They are regulation and the Zone of Proximal Development (ZPD).

As for the first concept, Lantolf and Thorne (2007) believe that learning, as a social process, develops through three stages—object-regulation, other-regulation, and self-regulation. At the object-regulation stage, learners’ behaviors are monitored by objects in their environment. At the other-regulation stage, help and guidance from others, such as teachers and peers, are provided to enable learners to progress. At the self-regulation stage, learners are able to act autonomously. According to Sociocultural Theory, CF provides learners with opportunities to self-repair their errors and ultimately to acquire the target features. CF functions as a form of mediation that assists learners to reach the last stage of self-regulation.

The second concept, Zone of Proximal Development (ZPD), is “the distance between the actual developmental level as determined by independent problem solving and the level of potential development as determined through problem solving under adult guidance or in collaboration with more capable peers” (Vygotsky, 1978, p.86). Within ZPD, learners can be pushed to a higher level of development and acquire what they are capable of acquiring with others’ assistance (Nassaji & Swain, 2000). From this perspective, CF provides negotiated assistance to prompt learners to progress from their
current level to a more sophisticated level of language development.

To sum up this section, different theoretical perspectives provide support for the practice of CF in language development. Theoretical accounts explain the facilitative role of CF from both cognitive and sociocultural aspects.

2.3.2 The Role of Metalinguistic Corrective Feedback in SLA

For almost three decades, researchers have shown growing interest in the research of corrective feedback. Based on their observational data from classrooms, early descriptive studies provided answers to questions such as what types of errors were corrected; when teachers corrected errors; how teachers provided correction (for example Allwright, 1988; Chaudron, 1977, 1988; Fanselow, 1977; Hendrickson, 1978). Recent descriptive studies addressed issues such as the different types of CF in communicative or content-based classrooms, the relationship between CF and learners’ uptake, and the relationship between learners’ error types and teachers’ CF types (for example Doughty, 1994; Han, 2001; Lyster & Ranta, 1997; Lyster, 1998; Seedhouse, 1999; Panova & Lyster, 2002; Sheen, 2004). Without doubt, corrective feedback is a key beneficial feature of student-teacher interaction. Numerous lab-based experimental studies (Carroll & Swain, 1993; Han, 2002; Mackey & Oliver, 2002) and classroom-based quasi-experimental studies (Doughty & Varela, 1998; Ellis, 2007; Ellis & Sheen, 2006; Lyster, 2004; Saito & Lyster, 2012; Sheen, 2007, 2010; Sheen & Ellis, 2011; Yang & Lyster, 2010) have provided supporting evidence for the efficacy of CF in learners’ interlanguage development. Statistically, the positive role of CF has also been confirmed by several
meta-analysis studies (for example Norris & Ortega, 2000; Russell & Spada, 2006; Mackey & Goo, 2007; Li, 2010; Lyster & Saito, 2010; Spada & Tomita, 2010).

2.3.2.1 The Role of MCF in Comparison with Other CF Types

There are established taxonomies of CF types among CF researchers and language teachers (Sheen, 2011). CF can be categorized into input-providing CF and output-prompting CF in terms of the provision of correct forms. CF can also be categorized into explicit CF and implicit CF in terms of the degree of explicitness. Recasts are the most frequently practiced implicit and input-providing CF type by language teachers; they are also the most intensely studied CF type by researchers (Lyster, 1998; Sheen, 2004; Lyster & Mori, 2006). MCF is a typical example of explicit and output-prompting CF. Ellis et al. (2006) assert that MCF and recasts “constitute the best exemplars of explicit and implicit corrective feedback, as both are supported by the previous research” (p.365).

This point of view is in line with the research results of a large number of studies that focused on the comparison of differential effects of explicit and implicit CF types in the last decade. However, some CF researchers (Lyster, 2002; Li, 2010) have challenged the explicit/implicit categorization. Lyster (2002) questioned the reliability of comparing the efficacies of CF in terms of explicitness. He asserts that it is pushed self-repair rather than explicitness that contributes to the language development. Likewise, Swain (1985, 1995, 2005) and De Bot (1996) hold that learners should be pushed to produce and modify their output for further language development to occur.
MCF, an explicit type of CF, clearly signals to the learner that an error has been made with the use of metalinguistic comments. It is also an output-prompting type of CF that provides learners with opportunities to detect their errors and modify their erroneous output. MCF requires learners to retrieve the target forms themselves and engages learners in a deeper level of processing. Learners are prompted to re-analyze what they have already learned at some level and restructure their interlanguage (Lyster, 2002). Compared to input-providing CF types that restrict learners’ opportunity of self-correction, MCF is a valuable opportunity for EFL learners, whose language learning almost entirely takes place in the classroom, to test the hypotheses about the use of linguistic forms and to be able to achieve higher levels of accuracy (Swain, 1995). According to De Bot (1996), EFL learners benefit more from being pushed to retrieve target language forms than from merely hearing the correct forms in the input, because retrieval and subsequent production stimulate the development of connections in memory.

2.3.2.2 Empirical and Experimental MCF Studies

The superiority of MCF over input-providing and implicit CF such as recasts is further supported by studies that examined the role of MCF in language learning, particularly those comparing differential effects of MCF and other input-providing CF (Lyster & Ranta, 1997; Ellis et al, 2006; Ellis, 2007; Lei, 2008; Sheen, 2011). One example is Lyster and Ranta’s (1997) observational study of French classroom audio-recordings. They investigated the effectiveness of types of corrective feedback in a
French immersion classroom. Four teachers provided corrective feedback on learner errors in speech production in 14 subject-matter lessons and 13 French language art lessons. The researchers classified feedback into six types: explicit correction, recasts, clarification requests, metalinguistic feedback, elicitation, and repetition. The results revealed that the most frequent type of feedback was the recast. However, MCF was found to be more effective on encouraging learners to generate repair. Lyster and Ranta concluded that CF can lead to learner uptake when there is “negotiation of form, the provision of corrective feedback that encourages self-repair involving accuracy and precision” (p.42) and when cues are given to make students aware of the necessity of repairing ill-formed utterances.

In the second example Ellis, Loewen, and Erlam’s (2006) study lends support for the efficacy of MCF in language learning as well. Ellis et al. compared the role of recasts and MCF in the acquisition of past tense –ed by ESL learners. The study is different from many other CF studies because it measured learners’ procedural knowledge in addition to declarative knowledge. 34 lower intermediate ESL learners from a private language school in New Zealand participated in this study. 77% of the participants were from East Asian countries. They were put into the recast group, the MCF group, and the control group. The participants took the pretest five days prior to the two instructional treatments. The posttest was completed the second day after the second instructional treatment and the delayed posttest was completed 12 days later. Each test involved three sections--untimed grammaticality judgment test, metalinguistic knowledge test, and oral imitation test. The first two were designed to measure the declarative knowledge and the last was
to measure the procedural knowledge. The instructional treatments took the forms of two different half-hour communicative tasks of story retelling and picture depicting. One of the researchers provided either recasts or MCF to the two experimental groups when the participants made past tense errors. Each experimental group engaged in a total of two half-hour communicative tasks during which they received Corrective Feedback. ANCOVAs were performed and the results indicated that (1) CF contributed to the development of both declarative knowledge and the procedural knowledge; (2) MCF was found to be more effective than recasts.

The third example is Ellis’ (2007) attempt to compare the differential effects of recasts and MCF on the acquisition of two grammatical structures in a quasi-experimental study. This study, similar to the Ellis et al. (2006) study, tapped not only declarative knowledge but also procedural knowledge. The two target structures under investigation were comparative –er and past tense –ed. Ellis listed two principal conditions for the choice of target structures: learnable and distinguishable in terms of learning difficulty. Drawing on Pienemann’s (1998) Processability Theory, which holds grammatical structures are acquired in an order reflecting the hierarchical processing operations, Ellis hypothesizes that past tense –ed will emerge before the comparative –er. 34 low-intermediate level ESL learners in a private language school in New Zealand were divided into three groups—namely, the recast group, the MCF group, and the control group. The researcher utilized an oral imitation test to measure procedural knowledge and an untimed grammaticality judgment test and a metalinguistic knowledge test to measure declarative knowledge. The three groups were asked to take a pretest, a posttest, and a
short delayed posttest two weeks after the posttest. The descriptive statistic results showed that the grammatical scores of the metalinguistic group for both past tense –\textit{ed} and the comparative –\textit{er} had increased from the pretest to the posttest in the oral imitation test. Split plot ANOVA (SPANOVA) was employed to find out answers to whether recasts and MCF had differential effects on the two target structures. The results indicated that there was no significant difference for recasts on the acquisition of the two target structures; however, MCF was found to have a greater effect on comparative –\textit{er} than on past tense –\textit{ed}. MCF was also found to have a great impact on the development of procedural knowledge of the two target forms—“a greater impact initially on comparative but a greater delayed impact on past tense ‘-ed’ ” (p.359). Ellis posited that MCF was more salient than recasts to assist learners’ acquisition and recasts need to be intensive and salient to produce limited effects. He also concluded that MCF was able to “freshen up” learners’ declarative knowledge in addition to its positive impact on procedural knowledge.

Another instance is Lei’s (2008) comparison of the efficacy of two CF types on Chinese college students’ acquisition of third person singular morpheme. She equally divided 54 non-English-majors from a college in south China into three groups—the recast group, the MCF group, and the control group. The treatment groups received two sessions of corrective feedback treatment from the instructor during the week after the pretest and before the posttest. Each session lasted 20 minutes. Lei employed picture description provided with key words and expressions as measurement. The posttest result showed a distinct group difference in accuracy rates at a \textit{p}-value of 0.004. The result
indicated that the metalinguistic feedback group outperformed the recast and the control groups with an accuracy mean score of 0.75. The result of the delayed posttest, which was conducted one week after the posttest, did not show a distinct group difference though the MCF group scored higher than the other two groups. Lei’s short-term research proved the effectiveness of MCF in reducing third person singular morpheme errors.

Finally, Sheen’s (2011) study also provides support for the beneficial role of MCF in SLA. This quasi-experimental study addressed the differential effects of recasts and MCF on the acquisition of the English definite and indefinite articles. 99 intermediate level students from a community college on the east coast of the United State participated in the study. They had stayed in the US for 1 to 5 years before the study and were native speakers of Chinese, Japanese, Korean, Spanish, Polish, Russian, and Turkish. There were three groups in the study—the recast group, the MCF group, and the control group. Recasts were operationalized as reformulation of a student’s incorrect utterance. MCF was operationalized as the provision of the correct form followed by metalinguistic information. This was different from the commonly-accepted definition of MCF by other researchers (Ellis et al, 2006; Lyster, 2004; Lyster and Ranta, 1997), who operated MCF as the provision of metalinguistic information to learners but withholding correct forms. The CF treatment took the form of a narrative task. In each of the two treatment sessions, the students were asked to read a story and then to retell it to the class. The stories served as stimuli to elicit article errors from the students. The two treatment sessions were audio-recorded. The testing instruments involved the pretest, the posttest, and the delayed posttest, each of which included a speeded dictation test, a writing test, and an error
correction test. A short exit questionnaire was also administered after the delayed posttest to examine the students’ awareness of the focus of the tests and the CF treatments. A one-way ANOVA and two-way ANOVAs with post hoc multiple comparisons were performed to investigate whether CF has a significant effect on the learning outcome. The results showed that MCF group outperformed the recast and the control groups in both the posttest and the delayed posttest. Pearson’s Chi-square tests suggested that the MCF group were more likely to notice that the focus of the study was relating to grammar than the other two groups. Sheen asserted that MCF resulted in the acquisition of English articles whereas the recast did not.

In sum, the studies reviewed here point to the facilitative role of MCF in language learning, whether French or English, past tense –ed or comparative –er. Compared with other input-providing and implicit CF types such as recasts, MCF exhibited a greater impact on the development of both declarative and procedural knowledge in language learning.

2.3.2.3 Meta-analysis Studies of MCF

In addition to the experimental and quasi-experimental studies reviewed earlier, the facilitative role of MCF in language learning has also been demonstrated by many meta-analysis studies. Norris and Ortega first introduced meta-analysis method in SLA in 2000. 49 unique sample studies from 45 study reports published from 1980 to 1998 were included in Norris and Ortega’s meta-analysis study (2000). Cohen’s $d$ was calculated to estimate the effect size of the L2 instruction. Comparisons of studies examining the
efficacy of MCF and recasts were also involved in the research synthesis. The average effect size for recasts was $d=0.81$ and for MCF was $d=0.96$. Although both found a large effect size ($d \geq 0.8$), the meta-analysis result suggested a larger effect size for this output-prompting CF type—MCF.

16 out of 28 interaction studies from early 1990s through June 2006 were included in Mackey and Goo’s (2007) meta-analysis of the efficacy of CF. The mean $d$ effect size was calculated for three CF types—recasts, MCF, and negotiation. The mean effect size was 0.96 for recasts, 0.47 for MCF, and 0.52 for negotiation on immediate posttest. The mean effect size was respectively 1.69 for recasts, 1.21 for MCF, and 0.58 for negotiation for the 9 studies that administered a short delayed posttest. The meta-analysis result suggested the effectiveness of MCF retained over a short term.

Mackey and Goo (2007) postulate that CF is more effective when it is provided consistently and has a narrow focus on linguistic items than being provided arbitrarily to deal with whichever errors arise in learners’ linguistic output. They found a larger mean effect size in specific focus conditions than in the general focus conditions through the meta-analysis and research synthesis of CF studies. To date, theoretical claims and empirical findings indicate that CF may be more or less beneficial depending on characteristics of the targeted linguistic forms (for example, Long, 1996; Long, Inagaki, and Ortega, 1998; Mackey, Gass, and McDonough, 2000). Mackey and Goo also claim that interaction plays a strong facilitative role in the acquisition of both lexical and grammatical items. They point out that interaction is particularly more beneficial for lexis than for grammar in the short-term, but more facilitative and durable for grammar in the
long-term.

Lyster and Saito (2010) meta-analyzed 15 classroom-based studies published from 1980 to 2010 to investigate the pedagogical effectiveness of oral CF on language development. Following Lyster and Ranta’s (1997) original taxonomy of CF, they examined recasts, prompts, and explicit correction in the study. MCF, together with clarification requests, repetition of error, and elicitation, was classified under the general typology of prompts. Cohen’s $d$ was calculated to examine whether CF was effective in classroom settings and whether the effectiveness varied according to CF types. The statistical results suggested that CF had significant and durable effects on language learning. Prompts, including MCF, were found to be more effective than recasts in classroom settings.

The efficacy of MCF in language learning has also been consistently supported by Li’s (2010) meta-analysis of 22 published CF studies and 11 unpublished doctoral dissertations between 1988 and 2007. MCF was examined as one of the explicit CF types in this meta-analysis. Seeking to understand the effectiveness of CF in L2 learning and the interactive effectiveness of CF types in relation to such variables as research settings, research contexts, and treatment lengths, Li calculated effect size $d$ and $Q$-tests. The results of Cohen’s effect size $d$ showed that CF exerted a medium effect on L2 development. The $Q$-tests results suggested the decrease of effect was not significant. Overall, CF was found to play a facilitative role in improving L2 learners’ linguistic accuracy and the effect was maintained over time. The explicit CF types including MCF were more effective than the implicit CF over a short term and the effect of implicit
feedback was better maintained over a long term. Nevertheless, Li pointed out that the effect size of the long-term effect of MCF was not calculated here “simply because there were not sufficient related studies” (p.349). Given the shortage of MCF studies for a long period, the long-term effect of implicit CF may be interpreted differently. Additionally, the meta-analysis also suggested that the lab-based studies showed a larger effect than classroom-based studies; shorter treatments generated a larger effect size than longer treatments; studies conducted in foreign language contexts produced larger effect sizes than those in second language contexts.

Taken together, the meta-analysis studies provide statistical support for the positive role of MCF already evidenced in many empirical and experimental studies. The meta-analyses results point to the greater impact of MCF than other implicit CF types in language learning.

2.4 Summary

In this chapter, I first reviewed gender referent errors and third person singular morpheme –s errors as two serious and persistent errors for Chinese EFL learners. In terms of declarative knowledge, intermediate and even advanced Chinese EFL learners learned the target forms and developed factual knowledge. In terms of procedural knowledge, they are not able to use the target forms fluently and accurately in spontaneous utterances yet. Research points to controversial positions regarding the treatability of the two target errors under investigation in this study. Although devastating, gender referent errors can be reduced to some extent with proper
intervention (Dong & Jia, 2011; Li & Han, 2010). In contrast, if learners did not receive focused CF, they continued to commit serious third person singular morpheme -s errors despite the formal instruction and the content teaching instruction they received during the period of study (Chang & Ma, 2006; Garcia Mayor & Villarreal Olaizola, 2010) or despite the long stay in English context (Lardiere, 1998a, b).

Next, I reviewed the necessity of distinguishing declarative knowledge and procedural knowledge as well as the valid measurement of two knowledge types. Because the two knowledge types are processed differently, different measurements are in need of attention from researchers. It was proposed that the metalinguistic knowledge test and the elicited oral imitation test are respectively best to measure declarative knowledge and procedural knowledge (Bowles, 2011; Ellis, 2005; Erlam, 2006).

Furthermore, theories in relation to CF were also reviewed. The Interaction Hypothesis, Output Hypothesis, Noticing Hypothesis, Skill Acquisition Theory, and Counterbalance Hypothesis from the cognitive perspective as well as Sociocultural theory from a sociocultural perspective were presented in this chapter. All of these theories have provided theoretical support for the beneficial role of CF in language development.

Last but not least, research demonstrates a more facilitative role of MCF in the acquisition of linguistic features such as past tense –ed, comparative -er, and definite and indefinite articles than other implicit and input-providing CF in interlanguage development (Lyster & Ranta, 1997; Ellis et al, 2006; Ellis, 2007; Lei, 2008; Sheen, 2011). Learners benefit more from being pushed to produce modified output than being provided with correct forms. This is also evidenced by many meta-analysis studies (Norris &
Ortega, 2000; Mackey & Goo, 2007; Li, 2010; Lyster & Saito, 2010).

In light of the literature reviewed in this chapter, it is evident that there is a gap between EFL learners’ declarative knowledge and procedural knowledge of gender referents and the third person singular inflectional morpheme –s. Although MCF has been evidenced to play a positive role in language learning in general, the impact of variables such as target features, the operationalization of MCF, linguistic knowledge types, and research contexts has also emerged. These variables continue to set the agenda for CF research (see Ellis, 2007; Ellis, 2012; Ellis et al., 2006; Ellis et al., 2009; Sheen, 2011; Lyster et al., 2013). The research questions guiding the present study are concerned with the role of MCF in assisting the development of declarative knowledge and procedural knowledge of two target features in EFL classroom context over an extended period of time. These research questions are:

Q1. Does oral metalinguistic corrective feedback increase the ability of Chinese university English majors to accurately assign gender referents? If yes, does the effect remain over an extended period of time?

Q2. Does oral metalinguistic corrective feedback increase the ability of Chinese university English majors to accurately produce the third person singular morpheme –s? If yes, does the effect remain over an extended period of time?

Q3. Does oral metalinguistic corrective feedback have differential effects on the acquisition of gender referents and third person singular morpheme –s?
CHAPTER THREE

METHODOLOGY

To examine the efficacy of Metalinguistic Corrective Feedback (MCF) in classroom settings over an extended period, the study employed a quasi-experimental, pretest, posttest, and delayed posttest research design over 36 weeks. The research questions were investigated by comparing the effects of MCF on two linguistic forms—gender referents and third person singular morpheme -s in classroom contexts. Learners’ awareness of the target features was investigated through a questionnaire following the last intervention instruction, together with the immediate posttest. Learners’ declarative knowledge of two target features was assessed through a metalinguistic knowledge test. Learners’ procedural knowledge was assessed through an elicited oral imitation test. Two intact classes of 60 students were randomly assigned to the metalinguistic corrective feedback group, hereafter referred to as the MCF group, and the control group.

This chapter describes the methodology used in this study. It begins with a description of the research context and the participants. It proceeds with research procedure, followed by an introduction of the target features. The chapter also outlines instructional intervention and testing instruments. Description of the data coding and scoring procedures is then provided. The chapter ends with a brief introduction of data analysis.
3.1 Research Context

The study was conducted in the School of English Studies at a university of foreign languages in northeast China. The participating university is the only foreign languages university in Northeast China. The university used to admit entirely students of foreign languages nationwide. In the recent decade, the university started to offer programs such as computer science, business, fine arts, etc. In the School of English Studies in the university, many programs, including the program involved in this research, began to adopt content-based language instruction in 2007 as a teaching reform. Besides some language-driven courses, the curriculum included some content-based courses such as American History, Greek Mythology, British Geography and so on.

Two intact classes taking the Communicative Spoken English Course participated in the study. A meaning-based oral English class in which participants had more oral production opportunities than in other courses was a desirable instructional context for this study since research showed that more errors emerged in speaking than in writing. The class met once a week for 90 minutes over 17 weeks during the semester. The teaching approach adopted in the class was communicative with an emphasis to develop students’ communicative language skills in English. The texts were mainly communication-based. There was a thematic focus in every unit. Table 3.1 lists the thematic foci in each unit over the semester.
Table 3.1 Thematic Foci of Each Unit for Class Intervention

<table>
<thead>
<tr>
<th>Unit</th>
<th>Thematic foci</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Campus life</td>
</tr>
<tr>
<td>2</td>
<td>Hobbies and interests</td>
</tr>
<tr>
<td>3</td>
<td>Movies and TV programs</td>
</tr>
<tr>
<td>4</td>
<td>Music</td>
</tr>
<tr>
<td>5</td>
<td>Computers and the internet</td>
</tr>
<tr>
<td>6</td>
<td>Sports and health</td>
</tr>
<tr>
<td>7</td>
<td>Weekends and holidays</td>
</tr>
<tr>
<td>8</td>
<td>Customs, festivals and celebrations</td>
</tr>
<tr>
<td>9</td>
<td>Sightseeing</td>
</tr>
<tr>
<td>10</td>
<td>Love and friendship</td>
</tr>
<tr>
<td>11</td>
<td>Fashion</td>
</tr>
<tr>
<td>12</td>
<td>Family life</td>
</tr>
<tr>
<td>13</td>
<td>Further study home and abroad</td>
</tr>
<tr>
<td>14</td>
<td>Jobs and work experience</td>
</tr>
</tbody>
</table>

3.2 Research Participants

There were 60 participating students from two intact classes. The two classes were randomly assigned to the MCF group and the control group. The participants were first-year English major students in the program of advanced English translation and interpretation. They were all native speakers of Chinese. Their ages ranged from 17 to 20.
They had completed at least six years of formal English education in middle and high schools before entering the program. Their length of formal English learning varied according to their previous residence provinces. They were enrolled in the program of English Studies after passing the threshold score of the National College Entrance Exams in China. They were regarded as intermediate English learners compared to other non-English major students in Chinese universities. The following table summarizes the bio-data of the participating students.

Table 3.2 Bio-Data of the Participating Students

<table>
<thead>
<tr>
<th>Class</th>
<th>Treatment</th>
<th>Number of students</th>
<th>Number of male students</th>
<th>Number of female students</th>
<th>Average age</th>
<th>Average years of formal English learning</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Control</td>
<td>31</td>
<td>3</td>
<td>28</td>
<td>19</td>
<td>10</td>
</tr>
<tr>
<td>2</td>
<td>MCF</td>
<td>29</td>
<td>4</td>
<td>25</td>
<td>19</td>
<td>10</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>60</td>
<td>7</td>
<td>53</td>
<td>19</td>
<td>10</td>
</tr>
</tbody>
</table>

To ensure both research groups received the same instructions except for the provision of MCF, the researcher was the participant teacher teaching both classes. She is a certified college English teacher in China for over 10 years. She is a native speaker of Chinese and a proficient English speaker. She received her master’s degree in TESOL from a university in the USA and is a doctoral student in a Language Education program in an American university. She has language teaching experience at the college level in both China and the USA. She also taught English in immigrants’ language training
programs funded by the American government. She taught the Communicative Spoken English Course during the period of the study.

3.3 Research Procedure

The study took place in two intact classes taking the Communicative Oral English Course over 36 weeks. One group received metalinguistic corrective feedback on the target features. The other group was the control group, which did not receive any corrective feedback on the target features. The research schedule in Figure 3.1 provides an overview of the entire research procedure.
Figure 3.1 Overall research design

- **Week 1**: Consent Form
- **Week 2**: Pretest
- **Week 3-16**: Communicative form-focused tasks with MCF
- **Week 17**: Communicative form-focused tasks without MCF
- **Week 36**: Long-delayed Posttest
Both classes met for 90 minutes per week over 17 weeks. In Week 1, the learners of both groups were asked for consent to take part in the study. In Week 2, the participating students were asked to complete the pretests measuring their declarative knowledge and procedural knowledge of two target structures. From Week 3 to Week 16, both groups received the same instruction except for CF on two target features. The learners from the MCF group received MCF on two target features from the teacher. The control group did not receive any CF on two target features. In order to monitor the proper instructional treatment, both classes were audio-recorded between pretests and immediate posttests. The data from in-class recordings also provided information about the distribution of learners’ errors, MCF provided, and modified output generated from MCF. In Week 17, both groups were asked to complete immediate posttests similar to the pretest and to complete a short exit questionnaire. The short exit questionnaire was employed to examine the learners’ awareness of research foci -- target features and MCF. In Week 36, the learners took a long delayed posttest similar to the previous two tests.

3.4 Target Structures

Two target structures were chosen for this study—third person singular morpheme –s and third person singular personal gender referents. There were two criteria for the selection—learners’ proficiency level and error types. As reviewed in the previous chapter, both target features were learned but not mastered by the majority of Chinese college students. Learners frequently committed detrimental errors, particularly in their oral production. Many students had developed declarative knowledge of the target
features, but not the procedural knowledge yet. In terms of error types, third person singular morpheme –s is a morphosyntactical error and a local error, while third person gender referents are meaning-based lexical errors and global errors. These two target features challenged learners differently. The selection of target features from different categories contributed to the comparison of the differential effects of MCF in relation to linguistic targets.

3.5 Instructional Intervention

Both groups followed the same syllabus of the Communicative Oral English course during the 14-week instructional treatment. One group, however, received MCF of the target features embedded in communicative form-focused tasks in addition to the regular instruction. The other group received the same instruction, completed the same communicative form-focused tasks but did not receive MCF when target errors were made.

The MCF group received instruction that consisted of:

1. metalinguistic explanations: The teacher gave explicit metalinguistic explanations of the target features and a couple of additional structures to the class.
2. form-focused tasks: The teacher incorporated specially designed communicative form-focused tasks to elicit the target features in addition to regular communicative instruction.
3. MCF: The teacher immediately gave MCF once hearing an incorrect use of the target structures but withholding the correct forms. The teacher required uptake
with repair. If errors were not repaired, the teacher repeated MCF and resorted to the second or the third move as needed until uptake with repair appeared.

The control group received instruction that consists of:

1. metalinguistic explanations: The teacher gave metalinguistic explanations of some selected structures including the target structures to the control class.

2. form-focused tasks: The teacher incorporated the same form-focused tasks in addition to regular communicative instruction in the control class.

3. no MCF: The teacher did not provide any correction in response to the errors in use of the target structures, but the teacher gave random corrections to other non-target errors.

In order to help the learners proceduralize their declarative knowledge of the target features through repeated practice elicited from MCF, communicative form-focused tasks were designed in addition to regular instruction according to the following criteria:

1. the tasks were in line with the course syllabus;

2. the tasks were appropriate to the learners’ language proficiency level;

3. the tasks had to elicit the target structures in meaningful contexts.
<table>
<thead>
<tr>
<th>Unit</th>
<th>Activities</th>
<th>Target features</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Campus life</td>
<td>A group of three students interviewed one another about the daily routines of their roommates. Later, some students were selected to share their interview with the whole class.</td>
<td>Gender referents and third-person singular morpheme</td>
</tr>
<tr>
<td>2. Hobbies and interests</td>
<td>A group of three students worked together and told one another the hobbies and interests of their three friends.</td>
<td>Gender referents and third-person singular morpheme</td>
</tr>
<tr>
<td>3. Movies and TV programs</td>
<td>After doing research on their favorite living actor and actress, students were asked to talk about the likes and dislikes of the actor and the actress.</td>
<td>Gender referents and third-person singular morpheme</td>
</tr>
<tr>
<td>4. Music</td>
<td>Before class, students were asked to do research on their favorite living musicians. In class, students were asked to write down the musician’s information on a piece of paper in this way—Name: Jackie Chan; Nationality: Chinese; Hobbies: watching movies, auto racing, collecting china. The teacher collected the information paper and randomly asked students to pick one piece of paper and to introduce the musician according to the information paper to the class.</td>
<td>Gender referents and third-person singular morpheme</td>
</tr>
<tr>
<td>5. Computers and the internet</td>
<td>In a group of three, students first described one of their roommates’ typical day with the use of some high-tech gadgets such as smart phones, MP3, and computers, etc. The other students took notes. Later, some students were selected to report the high-tech gadget use to the class.</td>
<td>Gender referents and third-person singular morpheme</td>
</tr>
<tr>
<td>6. Sports and health</td>
<td>Students were asked to tell the class what a male and a female family member/friend/relative/roommate did regularly to remain fit and healthy.</td>
<td>Gender referents and third-person singular morpheme</td>
</tr>
<tr>
<td>7. Weekends and holidays</td>
<td>Students were asked to describe typical weekend activities of their roommates in a group of three.</td>
<td>Gender referents and third-person singular morpheme</td>
</tr>
<tr>
<td>Topic</td>
<td>Activity</td>
<td>Gender referents and third-person singular morpheme</td>
</tr>
<tr>
<td>-------------------------------------------</td>
<td>--------------------------------------------------------------------------</td>
<td>-----------------------------------------------------</td>
</tr>
<tr>
<td>8. Customs, festivals and celebrations</td>
<td>Students were asked to tell their favorite holiday and the holiday traditions he/she does in groups. Later, they reported to the class how their group members celebrated the holiday.</td>
<td></td>
</tr>
<tr>
<td>9. Sightseeing</td>
<td>After reading a passage “Nature-Based Travel Threatens Fragile Ecosystem”, 3 students worked in a group and exchanged their ideas on the topic. One student from each group was selected to report to the class the other two group members’ opinions. Students were asked to start the report in this way—Lucy agrees/disagrees with the author. She thinks….</td>
<td></td>
</tr>
<tr>
<td>10. Love and friendship</td>
<td>After watching a video of four celebrities’ opinions on the one-child policy, students were asked to retell each celebrity’s opinion.</td>
<td></td>
</tr>
<tr>
<td>11. Fashion</td>
<td>Students were asked to talk about the fashion habits of one of their roommates.</td>
<td></td>
</tr>
<tr>
<td>12. Family life</td>
<td>After reading a story entitled “My Big Family”, students were asked to retell the story.</td>
<td></td>
</tr>
<tr>
<td>13. Further study home and abroad</td>
<td>In a group of three, students interviewed one another to find out whether they preferred to do further study at home or abroad and also why. After the group work, some students were asked to report their interview to the class.</td>
<td></td>
</tr>
<tr>
<td>14. Jobs and work experience</td>
<td>In a group of three, students interviewed one another to find out what their dream jobs were and why their dream jobs attracted them. After the group work, some students were asked to report their interview to the class.</td>
<td></td>
</tr>
</tbody>
</table>

MCF was operationalized as the “seamless” and “succinct” provision of explicit metalinguistic comments on target errors but withholding correct forms, following Reed’s (2012) operationalization of metacognitive CF in her study. This
operationalization was in line with the majority of CF researchers’ operationalization of MCF (Lyster, 2004; Ellis et al, 2006; Ellis, 2007; Yang & Lyster, 2010), but different from Sheen’s (2007, 2011), which was operationalized as the provision of the correct form in addition to metalinguistic information. When the teacher heard a target error in the learners’ oral production, the teacher provided immediate and consistent metalinguistic comments to the learners as in the following example.

For example:

Student: My mom always cook for us.

Teacher: Inflectional morpheme in simple present tense.

Student: My mom always cooks for us.

In order to familiarize the learners from the experimental group with metalinguistic terms of the target features, the teacher gave metalinguistic explanations of the target features in class before the first instructional intervention began. The consensus metalinguistic term of gender referents could be replaced with other mutually accepted names between the teacher and the students, for instance, gender, pronouns, or third person pronouns for the sake of mutual understanding and time saving in class.

Accordingly, third-person singular morpheme –s could be replaced with third-person –s, simple present tense, -s morpheme. To blur the research focus of two target features, a series of other frequently occurring errors including articles, plurals, past tense, real conditional and unreal conditional were reviewed simultaneously.

Instructional treatment sessions between pretests and posttests in both classes were recorded digitally and reviewed to confirm that all instructions were carried out as
planned. Learners’ errors, MCF provided, and uptake with repair were also calculated. The researcher and the other teacher teaching the same course coded the data to identify MCF on the target structures provided to the experimental group and confirm no CF on the target structures was provided to the control group. The data also rendered practical information of learners’ development on two target structures.

### 3.6 Testing Instruments

The effect of MCF on the acquisition of the target structures was assessed by collecting quantitative data from pretests, immediate posttests, and long delayed posttests. Each testing session was composed of an elicited oral imitation test following Ellis (2005), designed to measure learners’ procedural knowledge of the target structures and a metalinguistic judgment test designed to measure learners’ declarative knowledge. Three versions of elicited oral imitation tests and three versions of metalinguistic tests are administered in the pretest, the posttest, and the delayed posttest. A short questionnaire was administered immediately after the intervention and the posttest to evaluate learners’ awareness of the research foci and the CF treatment received in class. Learners’ attitudes toward CF were also explored through the questionnaire.

#### 3.6.1 The Elicited Oral Imitation Test

The elicited oral imitation test has been evidenced as a reliable and valid measure of procedural knowledge (Bowles, 2011; Ellis, 2005; Ellis et al, 2006; Erlam, 2006; Munnich, Flynn, & Martohardjono, 1994). In this study, the test consisted of 20 belief
statements involving both grammatical and ungrammatical sentences. Four out of twenty sentences were distractors. Sixteen sentences contained two target structures, 8 for each. Among 8 sentences containing gender referents, 5 cases required the use of masculine gender referents and 5 cases required the use of feminine gender referents. That is to say a sentence may contain two cases in use of gender referents. In all 10 cases, half were correct sentences that did not require correction and half were incorrect sentences that required spontaneous correction. The design of belief statements on third-person singular morpheme –s was similar to gender referents. Among 8 sentences containing third-person singular morpheme –s, 5 cases were correct sentences and 5 were incorrect sentences.

The learners took the elicited oral imitation test in an audio-visual language lab in the participating university. Learners were first asked to listen to each of the recorded sentences. They were then required to indicate whether they agreed with the statements, disagreed, or were not sure on a multiple choice test paper distributed to them before the test. The purposes were two-fold -- to focus learners’ attention on meaning and to delay their rote memory of the statements they heard. Learners were later asked to repeat what they heard in correct English. Learners were given limited time to finish their opinion multiple-choice part and the repetition part before they moved to the next sentence. Their responses were audio-recorded. Ellis (2009) believes that the dual tasks of deciding their opinions and imitating it “are more likely to elicit procedural knowledge because it places fewer demands on control mechanisms” (p.350). In order to familiarize the learners with the elicited oral imitation test, the researcher guided learners through with 4 practice sentences before the real test began.
3.6.2 The Metalinguistic Knowledge Test

The metalinguistic knowledge test has been proved the best valid test to measure declarative knowledge in a battery of tests that measure both declarative knowledge and procedural knowledge (Elder, 2009; Ellis, 2005; Bowles, 2011). The test in this study consisted of two parts. The first part took the form of multiple-choice questions to measure learners’ metalanguage. Learners were asked to select the right choice to answer the questions. These questions included selecting the rule that explained the error in the ungrammatical sentence, identifying the sentence in use of named grammatical parts, and selecting the appropriate target feature in compulsory contexts. This part contained 6 items, 3 for each target structure. In the second part, there were 2 passages. The test required learners to locate and underline gender referents in one passage and third-person singular morpheme –s in the other passage. Each passage contained 7 cases of each target structure. This type of test, used by Alderson et al. (1997), Elder et al. (1999), Newman and White (1999), is a valid measure of receptive knowledge of grammatical metalanguage (Ellis, 2004). Ellis further states that “this receptive measure is to be preferred” because the test sidesteps the problem of test takers’ guessing and makes “guesswork unrewarding” (p. 269).

Different from many studies that employed multiple choice or metalinguistic judgment test as the only measure of declarative knowledge, this study utilized a metalinguistic knowledge test of different forms. The purpose was to measure the use of declarative knowledge in various forms. The corrective feedback strategy under study was metalinguistic corrective feedback. MCF is distinguished from other CF strategies in
that it intentionally calls learners’ attention to their erroneous use of target structures and raises their awareness of metalinguistic knowledge of target structures. It is likely that learners’ awareness of the target structures could be raised through the test of locating target features. However, this is consistent with the purpose of class intervention activities.

Learners took the metalinguistic knowledge test immediately after the elicited oral imitation test. The tests were administered in a consistent order throughout the study to avoid the metalinguistic knowledge test priming the research participants.

3.6.3 Questionnaire

A short questionnaire was administered after the immediate posttest. It was adapted from Sheen’s (2007) exit questionnaire and Yang’s (2008) short questionnaire. The questionnaire in the current study was composed of a closed multiple-choice question and two open-ended questions. The purposes were twofold -- to evaluate whether learners were aware of the focus of the tests and the corrective feedback treatment and to examine learners’ attitudes toward corrective feedback. The two variables have been discussed in some studies in CF which investigated the effectiveness of CF. The multiple-choice questions were:

1. Now that you have completed the tests, what do you think they were all about?
   a) They were practicing and testing writing.
   b) They were practicing and testing grammar.
   c) They were practicing and testing listening and speaking.
They were practicing and testing vocabulary.

The two open-ended questions are:

2. Do you like to be corrected while you are speaking in class? Why or why not?

3. Please write a couple of sentences saying what you think you learned from the class activities and the tests.

The test was written in English, but learners were allowed to provide their answers in Chinese. Learners were supposed to provide more details and feel more comfortable in that case.

The following figure illustrates the procedure of the tests.
Figure 3.2 Testing Procedure

- **MCF Group**
  - Week 2: The elicited oral imitation test (pretest)
  - Week 17: The elicited oral imitation test (posttest)
  - Week 36: The elicited oral imitation test (Long-delayed Posttest)

- **Control Group**
  - Week 2: The metalinguistic test (pretest)
  - Week 17: The metalinguistic test (posttest)
  - Week 36: The metalinguistic knowledge test (Long-delayed Posttest)

- Exit questionnaire
3. 7 Scoring Procedure and Inter-rater Reliability

Different coding and scoring procedures were developed for each of the measures. In order to increase coding reliability, a second rater was involved in the coding and scoring process. The second rater, who is a native speaker of Chinese and an English assistant professor in the participating university, coded 25% of the oral and written data. The researcher went over the coding guidelines with the second rater and trained him to make sure he understood the procedures. Any discrepancy in coding was reexamined and recoded by the researcher and a third person. Disagreements were negotiated to reach final agreement. The inter-rater reliability reached 95%.

Declarative Knowledge Measure

Declarative knowledge tests consisted of two types of questions: multiple choice and locate and underline the target structures. Each was scored 10 if a correct answer was provided and was scored 0 if an incorrect answer or no answer was provided. Scores were expressed as percentage correct, i.e., participating students’ scores were the percentage of correct answers out of a maximum number of 10.

Procedural Knowledge Measure

Procedural knowledge was measured via the elicited oral imitation test. Accuracy was evaluated according to whether correct target features were supplied where they were required (Ellis, 2005; Ellis et al, 2007; Yang, 2008). This method was known as “suppliance in obligatory context (SOC)” in second language acquisition (Mackey &
Gass, 2005, p.232). In this study, the accuracy was operationalized as the correct use of target features in appropriate contexts where target features were required.

To be specific, in the case of gender referents, the correct use of gender referents and their respective case forms in the required context was scored 10. If a learner first gave a wrong answer and later self-corrected, the first incorrect answer was scored because only the initial incorrect production can represent the intuitive use of the target structure. A score of 0 was also given in the following circumstances:

1. Using a singular pronoun of incorrect gender.

   For example, Mozart is famous in that she is a talented musician.

2. Using a gender referent in the incorrect case form.

   For example, Hillary Clinton is a successful leader and people love she.

3. Using a plural pronoun where a singular gender referent was required.

   For example, Hillary Clinton is a successful leader and people love them.

The similar criterion was applied to score third-person singular morpheme –s. The correct use of third-person morpheme –s in compulsory context was scored 10. The following circumstances were scored 0:

1. Using any tenses other than simple present tense.

   For instance, It is raining less in Spring than in summer in Dalian.

2. Using base form of the verb without adding –s morpheme in simple present tense when the subject is third-person singular.

   For instance, An apple a day provide us with enough vitamins.

3. Omitting the predicate verb in the sentence.
For instance, Hanhan (lives) in Beijing for easy access to publishers.

4. Using the infinitive or the gerund of the verb.

For instance, A negative teacher to bring/bringing negative influence on students.

A total percentage accuracy score was calculated.

3.8 Data Analysis

In-class recording was transcribed and coded by the researcher and another English instructor from the university. The participants’ errors were counted as group errors along with the instructor’s and peers’ provision of MCF and learners’ uptake with repair. Descriptive statistics were run to examine the frequencies of learners’ errors of target structures, MCF provided to learners, and repairs following MCF.

Quantitative data collected from metalinguistic knowledge tests and the elicited oral imitation tests were entered and analyzed through the SPSS 22 package. Repeated measures ANOVAs were employed to evaluate and compare the effects of MCF on the declarative and procedural knowledge development of two target features. Effect sizes of Cohen’s $d$ were calculated to investigate the differential effect of MCF. In this study, students’ performances from the tests were the dependent variables. MCF technique and target structures were independent variables.

Data from the exit questionnaire were analyzed qualitatively and quantitatively to examine learners’ awareness of the research focus. As discussed and suggested in some studies on CF (see Sheen, 2007; Sheen, 2011; Yang & Lyster, 2010), students’ awareness is a strong indication of their growing monitor in speech production of target features,
which suggests their development of procedural knowledge. Participants’ different degrees of awareness could have an impact on the effectiveness of MCF.
CHAPTER FOUR

RESULTS

This chapter presents analysis and results of the data collected throughout the research. It starts with analysis of classroom transcriptions of 14 communicative form-focused tasks in the treatment sessions. Errors in use of target structures, feedback, and students’ uptake with repair were quantified and compared to provide a holistic picture of students’ performances of two target structures over the semester. Next, analysis of results from the elicited oral imitation test and metalinguistic knowledge test on two target structures are reported. Descriptive and inferential statistics were computed through SPSS 22 to provide answers to three research questions. Last, qualitative and quantitative analyses of the data collected from the exit questionnaire are presented. The analyses offer complementary information of learners’ noticing of the target features and their attitudes towards CF as well as their reflections about the research.

4.1 Analysis of Classroom Instructional Audio Data

In-class instructional treatment implemented through 14 communicative form-focused tasks was audio-recorded by 3 digital audio-recorders. The three recorders were placed in the front, the middle, and the back of students’ seating areas in classroom. A total of 689 minutes of instructional intervention was recorded, 352 minutes for the MCF group and 337 minutes for the control group. The recordings were transcribed and coded by the researcher and another English teacher. The number of errors in use of target structures, feedback, and repair were analyzed and quantified separately for each target
structure. The data detailed the implementation of CF and learners’ progress of their communicative use of target structures.

4.1.1 Coding Categories and Principles

In coding and analyzing the data, these principles were followed:

1. Errors on gender referents were quantified if a. a singular pronoun of incorrect gender was used as in Example 4.1; b. a gender referent in the incorrect case form was used as in Example 4.2; c. a plural pronoun was used where a singular gender referent was required as in Example 4.3.

Example 4.1

S: He always gets up early in spring festival and helps her mother prepare food.

(his)

Example 4.2

S: She’s favorite color is blue. (her)

Example 4.3

S: He often suggests we learn some medical knowledge because many of their patients were ill because they took the wrong medicine. (his)

2. Errors on the third person singular morpheme -s were quantified if a. any tenses other than simple present tense were used in obligatory contexts as in Example 4.4; b. a base form of a verb without adding –s morpheme in simple present tense was used when the subject was third-person singular as in Example 4.5; c. the infinitive or the gerund of a verb was used where a third person singular morpheme –s in simple present tense was
required as in Example 4.6.

Example 4.4

S: She plants her own food. Every time she eats them, she felt very proud. (feels)

Example 4.5

S: When she does some gardening, she feels that she just goes back to nature and forget everything. (forgets)

Example 4.6

S: Of course, no child deserving to be born in poverty. (deserves)

3. Feedback moves not directed at the target structures were excluded from analysis.

Sometimes the teacher provided incidental feedback on meanings and other errors than those in use of target structures, such as phonological errors and lexical errors. Feedback directed at phonological, lexical, and grammatical errors other than the target structures was neither consistent nor focused. Such feedback was not analyzed in this study.

4. Feedback from the teacher and from the students was not distinguished. Occasionally when an erroneous production was made, the other students would immediately or simultaneously correct the speaker. In this study, feedback was coded and analyzed regardless of its source. Both teacher-initiated and student-initiated feedback was counted as instances of feedback.

5. Self-generated repair and other-generated repair were not distinguished. Now and then, repair came from fellow students instead of the student at whom the teacher’s feedback was directed. The current study did not draw a line between the two as long as learners’ erroneous productions were repaired successfully.
6. Repeated feedback directed at one single erroneous production was considered and counted only once. In some circumstances, students failed to mend their errors correctly and the teacher used several MCF moves to elicit students’ correction as shown in Example 4.7. The successful and unsuccessful attempts were analyzed only once because they were directed at the initial erroneous production.

Example 4.7

S: She will make a call to order her supper on weekend evening.

T: The third person morpheme. (MCF)

S: She will makes a call. (uptake)

T: Third person simple present tense. (MCF)

S: She makes a call to order her food. (uptake with repair)

4.1.2 Results of Classroom Instructional Recording

In order to increase further understanding of students’ process in learning the two target structures and their development of procedural knowledge through communicative form-focused tasks with MCF, students’ erroneous production in use of target structures, the teacher’s and peers’ feedback as well as students’ repair were quantified and compared. Table 4.1 displays the distribution of errors, feedback, and repair on gender referents and Table 4.2 displays the counterpart information on the third person morpheme – s.
Table 4.1 Overall Distributions of Errors, Feedback Moves, and Repair on Gender Referents

<table>
<thead>
<tr>
<th></th>
<th>Error</th>
<th>Feedback</th>
<th>Repair</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Group</td>
<td>Group</td>
<td>Group</td>
</tr>
<tr>
<td>Task 1</td>
<td>13</td>
<td>12</td>
<td>10</td>
</tr>
<tr>
<td>Task 2</td>
<td>14</td>
<td>16</td>
<td>12</td>
</tr>
<tr>
<td>Task 3</td>
<td>11</td>
<td>18</td>
<td>10</td>
</tr>
<tr>
<td>Task 4</td>
<td>12</td>
<td>9</td>
<td>12</td>
</tr>
<tr>
<td>Task 5</td>
<td>14</td>
<td>11</td>
<td>13</td>
</tr>
<tr>
<td>Task 6</td>
<td>16</td>
<td>19</td>
<td>15</td>
</tr>
<tr>
<td>Task 7</td>
<td>8</td>
<td>15</td>
<td>8</td>
</tr>
<tr>
<td>Task 8</td>
<td>10</td>
<td>10</td>
<td>9</td>
</tr>
<tr>
<td>Task 9</td>
<td>12</td>
<td>16</td>
<td>11</td>
</tr>
<tr>
<td>Task 10</td>
<td>9</td>
<td>13</td>
<td>9</td>
</tr>
<tr>
<td>Task 11</td>
<td>11</td>
<td>14</td>
<td>10</td>
</tr>
<tr>
<td>Task 12</td>
<td>13</td>
<td>17</td>
<td>13</td>
</tr>
<tr>
<td>Task 13</td>
<td>12</td>
<td>15</td>
<td>11</td>
</tr>
<tr>
<td>Task 14</td>
<td>9</td>
<td>13</td>
<td>9</td>
</tr>
<tr>
<td>Total</td>
<td>164</td>
<td>198</td>
<td>152</td>
</tr>
</tbody>
</table>
Table 4.2 Overall Distributions of Errors, Feedback Moves, Repair on Third Person Morpheme –s

<table>
<thead>
<tr>
<th></th>
<th>Error</th>
<th>Feedback</th>
<th>Repair</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MCF Group</td>
<td>Control Group</td>
<td>MCF Group</td>
</tr>
<tr>
<td>Task 1</td>
<td>11</td>
<td>16</td>
<td>8</td>
</tr>
<tr>
<td>Task 2</td>
<td>21</td>
<td>22</td>
<td>17</td>
</tr>
<tr>
<td>Task 3</td>
<td>18</td>
<td>20</td>
<td>15</td>
</tr>
<tr>
<td>Task 4</td>
<td>16</td>
<td>17</td>
<td>15</td>
</tr>
<tr>
<td>Task 5</td>
<td>25</td>
<td>23</td>
<td>22</td>
</tr>
<tr>
<td>Task 6</td>
<td>19</td>
<td>27</td>
<td>18</td>
</tr>
<tr>
<td>Task 7</td>
<td>17</td>
<td>18</td>
<td>15</td>
</tr>
<tr>
<td>Task 8</td>
<td>13</td>
<td>21</td>
<td>12</td>
</tr>
<tr>
<td>Task 9</td>
<td>12</td>
<td>18</td>
<td>11</td>
</tr>
<tr>
<td>Task 10</td>
<td>11</td>
<td>16</td>
<td>10</td>
</tr>
<tr>
<td>Task 11</td>
<td>22</td>
<td>25</td>
<td>21</td>
</tr>
<tr>
<td>Task 12</td>
<td>12</td>
<td>17</td>
<td>12</td>
</tr>
<tr>
<td>Task 13</td>
<td>16</td>
<td>19</td>
<td>14</td>
</tr>
<tr>
<td>Task 14</td>
<td>13</td>
<td>18</td>
<td>13</td>
</tr>
<tr>
<td>Total</td>
<td>226</td>
<td>277</td>
<td>203</td>
</tr>
</tbody>
</table>

As shown in the tables, the control group produced 34 more errors on gender referents and 51 more errors on the third person morpheme –s than the MCF group. A possible reason could be the effect of treatment on improved productions in use of target
structures for the MCF group. The MCF group benefited from the treatment and got more opportunities to practice language use through reformulating their erroneous productions. Although in classroom settings teachers tend to provide CF of various types in a mixed manner, these tables revealed a relatively consistent provision of MCF to the experimental group.

In line with a number of explicit CF studies, the explicit and output-prompting MCF also induced a high repair rate in the current study. Altogether 149 repairs followed 152 feedback moves (98%) on gender referents and 192 repairs followed 203 feedback moves (95%) on the third person morpheme –s for the MCF group. Obviously students repaired more percentage of errors on gender referents than on the third person morpheme –s.

The nature of gender referents as lexical items is possibly conducive to more noticing from students and hence led to more repairs. In addition to the didactic force of MCF, the communicative purpose of MCF was perceived simultaneously when MCF was directed at gender referent errors. To clear up confusions, learners repaired errors on gender referents not only for the sake of being correct linguistically, but also for the sake of being accurate communicatively.

In terms of MCF on the third person morpheme –s in a communicative context, learners may ignore the didactic purpose of MCF but only care about getting meanings across. Two factors can account for learners’ neglect of the didactic force of MCF on the third person -s. Firstly, the communicative focus or meaning focus of the class could have an impact on students’ attention to meanings rather than forms. The present study
was carried out in a communicative oral English class. The class was more meaning-focused rather than form-focused. Most class activities were meaning-oriented. Participating students in the communicative class could overlook correctness of forms in their oral production as well as the corrective force in the teacher’s MCF, given that they expressed their intended meanings and were understood. Secondly, the linguistic feature of third person \(-s\) could have an impact on learners’ heedlessness. Third person morpheme \(-s\) is a rule-based and form-based grammatical feature. Although it has an evident morphosyntactic function, the semantic function is rather obscure. To the Chinese EFL learners in the current study, morpheme \(-s\) appears redundant and abstract because it does not convey meanings. The omission or misuse of morpheme \(-s\) does not necessarily cause communication failures.

Based on the above analysis, CF directed at phonological and lexical errors can be more salient than CF directed at morphosyntactic errors. This may also explain why more repairs on gender referents were generated than on the third person morpheme \(-s\). Overall, repeated practice of target features embedded through repair was evidenced in the MCF group.

### 4.2 Analysis of the Tests Results

To answer the research questions, three statistical models were employed separately for the elicited oral imitation test and the metalinguistic knowledge test. One-way ANOVA was employed to determine whether there is a significant difference between the control group and the MCF group in the pretest. Repeated measures
ANOVA were used to evaluate two issues. The first issue concerns different performances of the MCF group and the control group on their use of each target structure across three testing periods. The second issue is the interaction effect of time and treatment conditions on learners’ accuracy scores. The between-group variables were treatment conditions (MCF and control). The within-group variables were the timing of tests (pretest, posttest, and long delayed posttest). P-values were calculated to detect whether group differences were statistically significant. Post hoc (pairwise) comparisons were conducted to analyze the within-group difference over three testing periods. In addition, effect sizes of Cohen’s $d$ were computed and compared to investigate the differential effects of MCF on learners’ development of declarative knowledge and procedural knowledge on two linguistic targets. All statistics tests were set as two tailed and $p$-value was set at .05.

4.2.1 Gender Referents

Learners’ accuracy scores from the elicited oral imitation test and the metalinguistic knowledge test on gender referents were analyzed through one-way ANOVA. The result showed no statistically significant group difference in the pretest in both elicited oral imitation test, $F(1, 58)=.19, p>.05$, and the metalinguistic knowledge test, $F(1, 58)=.45, p>.05$ as shown in Appendix X Tables 1 and 2. Learners’ scores were then analyzed through repeated measures ANOVA. Cohen’s effect size $d$ was also computed. The following section will first present the results from the elicited oral imitation test followed by the results from the metalinguistic knowledge test.
4.2.1.1 Elicited Oral Imitation Test Results

Descriptive statistics of mean accuracy scores and standard deviation on gender referents in elicited oral imitation test are presented in Table 4.3. Figure 4.1 provides a graphic representation of the mean scores of MCF group and the control group across the three tests. It is evident that both groups reduced their erroneous use of gender referents over time. The MCF group started at a lower mean score of 53.45 in the pretest but outperformed the control group in the posttest and the delayed posttest.

Table 4.3 Gender Referents: Group Means and Standard Deviation in the Elicited Oral Imitation Test

<table>
<thead>
<tr>
<th></th>
<th>Pretest</th>
<th>Posttest</th>
<th>Delayed posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
</tr>
<tr>
<td>MCF group (n=29)</td>
<td>53.45</td>
<td>14.46</td>
<td>69.31</td>
</tr>
<tr>
<td>Control group (n=31)</td>
<td>54.84</td>
<td>9.62</td>
<td>58.71</td>
</tr>
</tbody>
</table>
Table 4.4 illustrates the results from repeated measures ANOVA of the two groups across the three tests. The results suggested that the MCF group and the control group performed differently in the tests. A significant effect for MCF on learners’ accuracy scores was evidenced by $F(1, 58)=6.63, p < .05$. A statistically significant time effect on learners’ accuracy scores was also found $F(2, 116)=8.86, p < .05$. Additionally, there was a significant interaction effect between time and feedback treatment, $F(2, 116)=3.17, p \leq .05$, indicating that the two groups performed differently from each other at tests. A pairwise comparison revealed a significant difference between the pretest and the posttest $F(1, 58)=15.64, p < .05$, but not a significant difference between the posttest and the long delayed posttest $F(1, 58)=.014, p >.05$. 

Figure 4.1 Gender Referents: Group Means of Percentage Scores in the Elicited Oral Imitation Test
Table 4.4 Gender Referents: Repeated Measures ANOVA in the Elicited Oral Imitation Test Across the Three Testing Periods

<table>
<thead>
<tr>
<th>Source</th>
<th>Df</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between subjects</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CF</td>
<td>1</td>
<td>6.63</td>
<td>.013</td>
</tr>
<tr>
<td>Error</td>
<td>58</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Within subjects</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>2</td>
<td>8.86</td>
<td>.000</td>
</tr>
<tr>
<td>Time × CF</td>
<td>2</td>
<td>3.17</td>
<td>.046</td>
</tr>
<tr>
<td>Error</td>
<td>116</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Effect sizes in Cohen’s $d$ were calculated through dividing the mean difference by the pooled standard deviation of the two compared groups. Cohen (1992) suggested that an effect size was considered large if it was above 0.8, medium if above 0.5, and small if below 0.2. MCF exerted moderate effect sizes in both the posttest (Cohen’s $d = .69$) and the delayed posttest (Cohen’s $d = .58$). The effect size of MCF slightly dropped .11 from the posttest to the delayed posttest.

4.2.1.2 Metalinguistic Knowledge Test Results

Table 4.5 presents mean accuracy scores and standard deviation on gender referents in the metalinguistic knowledge test. Figure 4.2 graphically displays the mean scores of MCF group and the control group across the three tests. Both groups began at an accuracy score over 80% in the pretest and slightly improved over time. As shown, the
MCF group outperformed the control group in all three tests despite the fact that both groups reduced their erroneous use of gender referents over time.

Table 4.5 Gender Referents: Group Means and Standard Deviation in the Metalinguistic Knowledge Test

<table>
<thead>
<tr>
<th></th>
<th>Pretest</th>
<th>Posttest</th>
<th>Delayed posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
</tr>
<tr>
<td>MCF group ((n=29))</td>
<td>85.17</td>
<td>9.86</td>
<td>93.10</td>
</tr>
<tr>
<td>Control group ((n=31))</td>
<td>82.90</td>
<td>15.53</td>
<td>86.77</td>
</tr>
</tbody>
</table>

Figure 4.2 Gender Referents: Group Means of Percentage Scores in the Metalinguistic Knowledge Test
Repeated measures ANOVA in Table 4.6 showed a significant time effect, $F(2, 116)=5.38, p < .05$, and group effect, $F(1, 58)=4.37, p < .05$, but not a significant interaction effect between time and feedback, $F(2, 116)=1.81, p > .05$. This indicated that the two groups did not perform significantly different from each other over time. A pairwise comparison revealed a significant difference between the pretest and the posttest $F(1, 58)=4.58, p < .05$, but not a significant difference between the posttest and the long delayed posttest $F(1, 58)=.59, p > .05$.

Table 4.6 Gender Referents: Repeated Measures ANOVA in the Metalinguistic Knowledge Test Across the Three Testing Periods

<table>
<thead>
<tr>
<th>Source</th>
<th>$Df$</th>
<th>$F$</th>
<th>$P$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between subjects</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CF</td>
<td>1</td>
<td>4.37</td>
<td>.041</td>
</tr>
<tr>
<td>Error</td>
<td>58</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Within subjects</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>2</td>
<td>5.38</td>
<td>.006</td>
</tr>
<tr>
<td>Time × CF</td>
<td>2</td>
<td>1.81</td>
<td>.68</td>
</tr>
<tr>
<td>Error</td>
<td>116</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Effect sizes in Cohen’s $d$ indicated that MCF exhibited a relatively small effect size of .43 in the posttest and .33 in the delayed posttest compared to the effect sizes in the oral test. The effect sizes of MCF narrowly dropped .1 from the posttest to the long
4.2.2 Third Person Singular Morpheme –s

This section will focus on learners’ performances in use of third person –s morpheme in the elicited oral imitation test and the metalinguistic knowledge test. A One-way ANOVA detected no significant differences between the groups in either test format, $F(1, 58)=.24, p >.05$ in the elicited oral imitation test and $F(1, 58)=.49, p >.05$ in the metalinguistic knowledge test. Statistical results of repeated measures ANOVA and Cohen’s $d$ effect sizes will be reported in the order of the elicited oral imitation test first, followed by the results from the metalinguistic knowledge test.

4.2.2.1 Elicited Oral Imitation Test Results

Table 4.7 displays mean accuracy scores and standard deviation of learners’ use of third person singular morpheme –s in the elicited oral imitation test. Figure 4.3 presents a graphic representation of the mean scores of the MCF group and the control group across the three tests. Both groups started at low and quite close accuracy scores in the pretest (the MCF group at 35.86 and the control group at 33.55), but improved over time. The MCF group improved dramatically from the pretest to the posttest (from 35.86 to 71.72) and moderately from the posttest to the delayed posttest (from 71.72 to 76.55). The control group improved more from the pretest to the posttest (from 33.55 to 43.87) but less from the posttest to the delayed posttest (from 43.87 to 48.71). Again, the MCF group outscored the control group at the three tests.
Table 4.7 Third Person Singular Morpheme -s: Group Means and Standard Deviation in the Elicited Oral Imitation Test

<table>
<thead>
<tr>
<th></th>
<th>Pretest</th>
<th>Posttest</th>
<th>Delayed posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
</tr>
<tr>
<td>MCF group (n=29)</td>
<td>35.86</td>
<td>18.23</td>
<td>71.72</td>
</tr>
<tr>
<td>Control group (n=31)</td>
<td>33.55</td>
<td>18.72</td>
<td>43.87</td>
</tr>
</tbody>
</table>

Figure 4.3 Third Person Singular Morpheme –s: Group Means of Percentage Scores in the Elicited Oral Imitation Test
Results of repeated measures ANOVA (Table 4.8) demonstrated a very significant effect of time $F(2, 116)=49.18, p < .001$, of group, $F(1, 58)=49.53, p < .001$, as well as of time $\times$ group interaction, $F(2, 116)=11.99, p < .001$. A pairwise comparison revealed a significant difference between the pretest and the posttest $F(1, 58)=49.8, p < .001$, but not a significant difference between the posttest and the long delayed posttest $F(1, 58)=3.09, p > .05$.

Table 4.8 Third Person Singular Morpheme -s: Repeated Measures ANOVA in the Elicited Oral Imitation Test Across the Three Testing Periods

<table>
<thead>
<tr>
<th>Source</th>
<th>$Df$</th>
<th>$F$</th>
<th>$P$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between subjects</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CF</td>
<td>1</td>
<td>49.53</td>
<td>.000</td>
</tr>
<tr>
<td>Error</td>
<td>58</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Within subjects</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>2</td>
<td>49.18</td>
<td>.000</td>
</tr>
<tr>
<td>Time $\times$ CF</td>
<td>2</td>
<td>11.99</td>
<td>.000</td>
</tr>
<tr>
<td>Error</td>
<td>116</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Cohen’s $d$ indicated very large effect sizes in both the posttest and the delayed posttest. The effect of MCF remained approximately the same from $d = 1.69$ in the posttest to $d = 1.70$ in the delayed posttest.
### 4.2.2.2 Metalinguistic Knowledge Test Results

As illustrated in Table 4.9 and Figure 4.4, both the MCF group and the control group started at high accuracy scores in the pretest (the MCF group at 87.59 and the control group at 85.48) and slightly improved their accuracy in use of third person morpheme –s over time. The MCF group outscored the control group in all three tests again. Mean scores for the MCF group did not show significant change from the posttest to the delayed posttest (from 97.59 to 97.93). The control group slightly increased its mean scores from the pretest to the posttest and from the posttest to the delayed posttest (from 85.48 to 90.97 and 91.29).

#### Table 4.9 Third Person Singular Morpheme -s: Group Means and Standard Deviation in the Metalinguistic Knowledge Test

<table>
<thead>
<tr>
<th></th>
<th>Pretest</th>
<th>Posttest</th>
<th>Delayed posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
</tr>
<tr>
<td>MCF group (n=29)</td>
<td>87.59</td>
<td>10.57</td>
<td>97.59</td>
</tr>
<tr>
<td>Control group (n=31)</td>
<td>85.48</td>
<td>12.61</td>
<td>90.97</td>
</tr>
</tbody>
</table>
Repeated measures ANOVA results (Table 4.10) demonstrated a significant time effect, $F(2, 116)=13.95, p < .001$, and group effect, $F(1, 58)=7.18, p < .05$, but not a significant interaction effect between time and feedback, $F(2, 116)=1.14, p > .05$. A pairwise comparison revealed a significant difference in learners’ performance in metalinguistic knowledge test between the pretest and the posttest, $F(1, 58)=14.1, p < .05$, but not a significant difference between the posttest and the long delayed posttest $F(1, 58)=.48, p > .05$. 
Table 4.10 Third Person Singular Morpheme -s: Repeated Measures ANOVA in the Metalinguistic Knowledge Test Across the Three Testing Periods

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between subjects</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CF</td>
<td>1</td>
<td>7.18</td>
<td>.010</td>
</tr>
<tr>
<td>Error</td>
<td>58</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Within subjects</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>2</td>
<td>13.95</td>
<td>.000</td>
</tr>
<tr>
<td>Time × CF</td>
<td>2</td>
<td>1.14</td>
<td>.323</td>
</tr>
<tr>
<td>Error</td>
<td>116</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Effect sizes in Cohen’s $d$ indicated that MCF exhibited a moderate effect size of .59 in the posttest and a nearly large effect size of .74 in the delayed posttest. Interestingly, the effect size of MCF did not decrease rather increased .15 from the posttest to the long delayed posttest.

### 4.2.3 Summary of the Tests Results

This section summarizes research findings generated from statistical results with a view to answering the three research questions raised in Chapter 1.

Q1. Does oral metalinguistic corrective feedback increase the ability of Chinese university English majors to accurately assign gender referents? If yes, does the effect remain over an extended period of time?
The answer is affirmative. The MCF group increased their accurate use of gender referents in both the elicited oral imitation test and the metalinguistic knowledge test over the posttest and the long delayed posttest periods. Learners benefited more from MCF improving their procedural knowledge of gender referents than their declarative knowledge. The effect of MCF on gender referents maintained over an extended period of time, 20 weeks after the treatment, as demonstrated by a significant time effect.

Q2. Does oral metalinguistic corrective feedback increase the ability of Chinese university English majors to accurately produce the third person singular morpheme –s? If yes, does the effect remain over an extended period of time?

Again, the answer is affirmative. The MCF group increased their accurate use of the third person singular morpheme –s in both the elicited oral imitation test and the metalinguistic knowledge test over the posttest and the long delayed posttest periods. The effect of MCF on the third person singular morpheme –s maintained 20 weeks after the end of instructional treatment, as demonstrated by a very significant time effect.

Q3. Does oral metalinguistic corrective feedback have differential effects on the acquisition of gender referents and third person singular morpheme –s?

To answer this question, a close examination of Cohen’s d effect sizes of MCF on the two linguistic targets is needed in addition to the statistical results from repeated measures ANOVA. Table 4.11 summarizes the effect sizes of MCF on gender referents and third person singular morpheme –s in the elicited oral imitation test and the metalinguistic knowledge test over time. In terms of linguistic targets, MCF exerted larger effect sizes on the third person morpheme –s than on gender referents in both the
elicited oral imitation test and the metalinguistic knowledge test. A thorough examination also reveals that the effect size of MCF on gender referents dropped from the posttest to the long delayed posttest (from 0.69 to 0.58 for procedural knowledge and from 0.43 to 0.33 for declarative knowledge). On the contrary, the effect size of MCF on procedural knowledge for third person -s remained more or less the same from the posttest to the long delayed posttest (from 1.69 to 1.70) and slightly increased .15 from 0.59 to 0.74 for declarative knowledge. This suggests a slightly enhanced effect on the third person morpheme –s in the long term. Such a finding was evidenced in a number of meta-analysis studies (see Li, 2010; Mackey & Goo, 2007). In terms of test formats, MCF exerted larger effects on the elicited oral imitation test than on the metalinguistic knowledge test. In other words, MCF played a more facilitative role in Chinese college EFL learners’ development of procedural knowledge on gender referents and the third person morpheme -s than their development of declarative knowledge on the two target structures.

Table 4.11 Effects of MCF on Gender Referents and Third Person Singular Morpheme –s in Different Tests (in Cohen’s d effect size)

<table>
<thead>
<tr>
<th></th>
<th>Elicited oral imitation test</th>
<th>Metalinguistic knowledge test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Posttest</td>
<td>Delayed posttest</td>
</tr>
<tr>
<td>Gender referents</td>
<td>0.69</td>
<td>0.58</td>
</tr>
<tr>
<td>Third person morpheme –s</td>
<td>1.69</td>
<td>1.70</td>
</tr>
</tbody>
</table>
4.3 Analysis of the Exit Questionnaire

Learners completed the exit questionnaire right after the first posttest. The questionnaire was composed of three questions relating to learners’ attitudes toward corrective feedback and their awareness of the focus of the research. All 60 students (29 in the MCF class and 31 in the control class) answered the questions. The following analyses were carried out on the information gathered from students’ responses.

Question 1 (Now that you have completed the tests, what do you think they were all about?) was a multiple-choice question. There were four choices available for learners--a) they were practicing and testing writing; b) they were practicing and testing grammar; c) they were practicing and testing listening and speaking; and d) they were practicing and testing my vocabulary. No one chose a) writing. Altogether 22 learners chose b) grammar. Among them 6 learners were from the control class and the rest 16 learners were from the MCF class. Thirty-seven learners chose c) listening and speaking. Among them 24 learners were from the control class and 13 were from the MCF class. Only 1 learner from the control class chose d) vocabulary.

Question 2 (Do you like to be corrected while you are speaking in class? Why or why not?) asked about learners’ attitudes towards CF. The majority of learners, 52 out of 60, gave affirmative answers. 28 learners from the control class and 24 learners from the MCF class expressed their positive attitude toward CF. According to them, the chief reasons for their likes included that they could improve their English and that being corrected is an impressive way to notice and learn from their errors. Only 5 learners, 3 from the control class and 2 from the MCF class, expressed negative attitudes toward
corrective feedback. Three reasons accounted for their dislikes. First, they disliked being interrupted while talking. Second, being corrected in class was embarrassing. Third, being corrected in class could build their tension. 2 of the 5 learners who disliked the correction also indicated their preference of being corrected after their talk. 3 learners, exclusively from the MCF class, expressed an ambiguous attitude “it depends”. They claimed that a “proper” and “kind” way to give corrective feedback was acceptable.

Question 3 (Please write a couple of sentences saying what you think you learned from the class activity and the tests.) was aimed to elicit more information about learners’ awareness and understanding of the research. As it is an open-ended question, learners’ answers showed more variety. Six themes--grammar, listening, speaking, mistakes and accuracy, memory, and practice--were analyzed and categorized in light of learners’ responses. 26 learners, 10 from the control class and 16 from the experimental class, stated that they improved or needed to improve their grammar use. Interestingly, most learners who detected the grammar focus of the tests were also able to tell explicitly both or one of the two target forms in this research, with 5 more reports of 3rd person –s morpheme than the reports of gender referents. 13 learners, 9 from the control class and 4 from the experimental class, attached importance to listening. 16 learners, 6 from the control and 10 from the experimental class, emphasized the importance of speaking in improving English. 6 learners, 3 from each group, claimed mistakes and accuracy hindered their better performances. 9 learners, 3 from the control class and 6 from the experimental class, mentioned the effect of memory in completing the tasks in the tests. 8 learners, 3 from the control and 5 from the experimental class, expressed their need for
more practice. Although diverse research participants’ answers appeared, they provided additional information and support for their noticing of the research foci. Learners’ attention to the research foci suggests learners’ noticing of the target structures and MCF embedded in class activities. Noticing has been discussed as an indispensable element for learners’ improved accuracy in their later oral production. MCF contributes to raising learners’ awareness and elevating learners to a higher level of understanding and incorporating the metalanguage of target structures in their communication. It is thus feasible for learners to convert declarative knowledge to procedural knowledge. In other words, noticing of errors and MCF is crucial to EFL learners’ proceduralization of their declarative knowledge. More reported noticing from the MCF group than the control group in this study may provide an explanation for the better performance of the MCF group in the posttest and the long delayed posttest from another perspective.
CHAPTER FIVE

DISCUSSION AND CONCLUSIONS

This chapter provides a summary of the research findings. Potential factors such as explicitness of MCF and the nature of linguistic features that may have influenced the efficacy of MCF are reviewed and discussed against those found in other previous studies. In what follows, pedagogical implications of research findings and limitations of this study are described. This chapter concludes with suggestions for future research.

5.1 Summary of Research Findings

This study explored differential effects of metalinguistic corrective feedback (MCF) on the development of declarative knowledge and procedural knowledge of two linguistic features, i.e., gender referents and third person morpheme -s. It was hypothesized that learners receiving MCF were advantaged in their development of declarative knowledge and procedural knowledge of both linguistic features over learners receiving no MCF. Two meaning-based Communicative Oral English classes in mainland China were chosen for the study, with English gender referents and third person morpheme -s as linguistic targets. A total of 60 intermediate-level Chinese college EFL learners from two intact classes were randomly assigned to the MCF group and the control group. Using a pretest-posttest-delayed posttest design, the participants in each group completed 14 communicative form-focused tasks in regular classes over the semester. The MCF group received MCF on their nontarget-like oral production during treatment while the control group did not. Both groups took the metalinguistic knowledge
test measuring declarative knowledge and the elicited oral imitation test measuring procedural knowledge on the two linguistic targets respectively over three testing periods. They also completed an exit questionnaire immediately after the first posttest. The following summarizes the research findings.

**Table 5.1 Summary of Group Means Over Time**

<table>
<thead>
<tr>
<th></th>
<th><strong>Elicited Oral Imitation Test</strong></th>
<th><strong>Metalinguistic Knowledge Test</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Gender Pronoun</td>
<td>Third-person -s</td>
</tr>
<tr>
<td><strong>Gender Pronoun</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pretest MCF Group</td>
<td>53.45</td>
<td>35.86</td>
</tr>
<tr>
<td>Pretest Control Group</td>
<td>54.84</td>
<td>33.55</td>
</tr>
<tr>
<td>Posttest MCF Group</td>
<td>69.31</td>
<td>71.72</td>
</tr>
<tr>
<td>Posttest Control Group</td>
<td>58.71</td>
<td>43.87</td>
</tr>
<tr>
<td>Long Delayed Posttest MCF Group</td>
<td>69.66</td>
<td>76.55</td>
</tr>
<tr>
<td>Long Delayed Posttest Control Group</td>
<td>59.03</td>
<td>48.71</td>
</tr>
</tbody>
</table>

Table 5.1 provides an overview of research participants’ accuracy scores on two target structures over three testing periods. The MCF group outscored the control group in 11 out of 12 tests over the three testing periods. The only exception was in the elicited oral imitation test on gender referents in the pretest period. The mean score of the MCF
group (53.45) was slightly lower than the control group (54.84) by 1.4. In terms of gender referents, both groups increased their accuracy scores in the elicited oral imitation test and the metalinguistic knowledge test over the posttest and the long delayed posttest periods. Statistically significant group differences were also detected in both test formats over the posttest and the long delayed posttest periods, but not in the pretest. Learners benefited more from MCF in improving their procedural knowledge of gender referents than their declarative knowledge, demonstrated by a higher Cohen’s $d$ effect size ($d=0.69$ vs. $d=0.43$ in the posttest and $d=0.58$ vs. $d=0.33$ in the delayed posttest). The effect of MCF on gender referents persisted over an extended period of time for both groups. In terms of third person morpheme –s, the MCF group and the control group manifested all the research findings observed on gender referents, i.e. the MCF group performed significantly better than the control group in both posttest and delayed posttest. Two noteworthy facts should not be neglected. The first fact was that both groups improved more significantly from the pretest to the posttest on the third person –s than on gender referents. The second fact was that their improved performances were not only retained in the long delayed posttest but also enhanced to a different degree over an extended period of time. All in all, the finding is consistent with most studies of MCF, which evidenced the positive role of MCF in interlanguage development as well as its persisting effect over time (Ellis et al, 2006; Ellis, 2007; Lei, 2008; Li, 2010; Lyster & Ranta, 1997; Mackey & Goo, 2007; Sheen, 2011).

A more careful comparison of effect sizes of MCF on the two linguistic targets revealed that MCF exhibited different effects on the two linguistic targets. Larger effect
sizes on third person morpheme –s than on gender referents were found in both the elicited oral imitation test and the metalinguistic knowledge test. Another interesting finding was that test formats mediated the effects of MCF. It was evidenced that MCF exerted a larger effect on the elicited oral imitation test than on the metalinguistic knowledge test regardless of linguistic targets. This finding aligned itself with the majority of the empirical studies (Ellis et al., 2006; Li, 2014; Loewen & Nabei, 2007) and some meta-analysis studies (Li, 2010; Lyster & Saito, 2010), all of which investigated the effect of CF through tests tapping into learners’ declarative knowledge and procedural knowledge. However, this finding contradicts the finding from Yang & Lyster’s study (2010). They detected an overall larger effect size in written tests tapping into declarative knowledge than in oral tests tapping into procedural knowledge.

To conclude, three major research findings can be summarized and highlighted in relation to the findings obtained from other previous research. First, MCF played a facilitative role in learners’ acquisition of gender referents and the third person morpheme –s and the effects persisted over time. This finding showed congruence with previous studies (Carroll & Swain, 1993; Ellis et al, 2006; Ellis, 2007; Lei, 2008; Sheen, 2011), which evidenced a facilitative role of MCF in interlanguage development. Second, the effect of MCF was constrained by linguistic targets. MCF exerted a more facilitative role in the acquisition of the third person morpheme –s than gender referents. Previous research investigating the differential roles of CF on more than one linguistic target evidenced the superior role of explicit CF including MCF on certain linguistic targets over others (Ellis, 2007; Li, 2014; Yang & Lyster, 2010). Third, test formats mediated the
effect of MCF. In the aspect of knowledge types measured by different test formats, MCF played a more conducive role in learners’ development of procedural knowledge than the development of declarative knowledge regardless of target structures. Some previous studies (Ellis et al., 2006; Li, 2010, 2014; Loewen & Nabei, 2007; Lyster & Saito, 2010) produced similar findings.

5.2 Efficacy of Metalinguistic Corrective Feedback

The efficacy of MCF has been evidenced in a variety of learning contexts. Factors mediating its efficacy include internal factors such as innate features of MCF as well as external factors such as linguistic targets (Wang & Liu, 2013). This section first reviews the internal factors regarding how MCF is categorized in terms of its two most important features — salience and output-based orientation. Linguistic targets as external factors are discussed later.

5.2.1 Explicitness and Noticeability of MCF

In light of established taxonomies of CF types in the field, MCF is categorized as the most explicit of the CF types for its degree of explicitness. Unlike implicit feedback, MCF provides negative evidence instead of positive evidence to deliberately call learners’ attention to errors. The provision of metalinguistic comments clearly signals the existence of errors in learners’ utterances. In addition, it requires a break from the natural flow of communication and allows learners to focus their attention on erroneous output. Learners can effortlessly perceive the didactic purpose of MCF.
The in-class recording of communicative form-focused tasks revealed that the majority of learners were aware of their errors upon receiving MCF although some failed to provide uptake with successful repair as shown in Example 5.1.

Example 5.1.

S: Amy’s father always **cooking** for her and her mom.

T: Third person present.

S: Amy’s father always **cook** for her.

T: -s morpheme.

S: cooks for her.

Another evidence manifesting learners’ noticing of MCF is the phenomenon that learners noticed MCF that was addressed to other learners as presented in the following example.

Example 5.2

S1: My father doesn’t sleep well. He often **drank** after midnight.

T: Simple present tense.

S2: Drinks.

S1: Drinks. Drinks after midnight.

Data from the exit questionnaire also provided support for learners’ noticing of the explicit MCF. In answering Question 1 and Question 3, both examining learners’ noticing of research foci, more learners from the MCF group reported noticing the research foci than from the control group. There were respectively 22 reports of research foci in Question 1 and 26 reports in Question 3. Learners reported their noticing as follows:

“I learn to take notice of the grammar of the sentences, not just the meaning.”
“The test attaches much importance to grammar…I guess in the future in order to be a good English speaker, I should not only focus on language fluency, but also the correctedness.”

“We should pay attention to our grammar as well….”

Their answers to the questions also showed that most learners who noticed the grammar focus could even identify both or one of the two target forms:

“I should pay special attention to gender and tense. It’s not difficult but easy to be ignored in oral English.”

“We should always be careful using the gender pronouns and third-person (-s).”

“Pay attention to gender pronouns and improve my listening.”

“I should pay more attention to third person –s in my speech……”

The second question in the exit questionnaire (Do you like to be corrected while you are speaking in class? Why or why not?) inquired about learners’ attitudes towards CF. When the MCF group explained their reasons for their preferences for MCF, some interesting answers concerning noticeability also emerged:

“When I was corrected, I noticed my mistakes in speaking.”

“We need to realize our errors with the help of others.”

“(CF) makes the mistake more impressive to me.”

“I will have a deep impression of mistakes……I will be aware of them”

“It [being corrected] is an embarrassing experience for me. But considering the strong impression of my mistakes, it’s worth being corrected.”

Learners’ noticing of research foci was also evidenced in some other CF studies.
For example, Sheen (2007) investigated the differential effect of recasts and MCF on the acquisition of English articles. She found that learners receiving MCF reported more noticing in the exit questionnaire than learners receiving recasts or no CF. The research findings revealed that the MCF group outperformed the recast group and the control group; additionally, the recast group and the control group did not show a significant difference.

Likewise, the correlation between noticing and efficacy of MCF was observed in Yang and Lyster’s study (2010). They explored the role of prompts and recasts on Chinese EFL learners’ acquisition of past tense. More learners from the explicit CF group, i.e. prompts, reported noticing in their exist questionnaire than from the recasts group. The improved accuracy was evidenced in all eight measures for the prompt group, 4 for the recast group, and 3 for the control group.

Other studies also evidenced the positive role of MCF. However, they did not include the examination of learners’ noticing in their studies. The two studies together with the current one provide support for Gass & Selinker’s claim that learners’ noticing may not come naturally and “may require some pedagogical training” (2008, p.381). MCF can trigger learners’ noticing and perform this noticing function well.

Besides MCF’s innate feature of being salient, another factor also accounts for its noticeability—how MCF is implemented. In this study, MCF was provided to learners consistently and repeatedly on two focused linguistic features in all of the 14 instructional treatment sessions. The intensive provision of MCF over time helped to enhance the salience of MCF and to raise learners’ awareness. Subsequently, it prompted more
student-generated CF. This was evidenced by increasing peer feedback over time, captured in the in-class recordings. The data showed that learners mostly relied on the teacher to provide MCF on the target errors in the first few weeks of the treatment sessions. With time, learners noticed the intensively treated errors and started to join the teacher to provide peer feedback. As reviewed in Chapter 2, the effect and the salience of CF are constrained by the operationalization of CF. A consistent and focused provision of MCF is likely to be more conducive to language development than an incidental and unfocused provision.

Schmidt (1990, 1994, 1995, 2001) claims in the Noticing Hypothesis that noticing is crucial for language development. According to him, unlike L1, SLA is a conscious process and the provision of CF can promote learners’ noticing of the gap between their interlanguage and the target language. Schmidt (2010) further details the two levels of awareness—notice and understanding. He explains “‘noticing’ as a technical term limited to the conscious registration of attended specific instances of language, and ‘understanding’, a higher level of awareness that includes generalizations across instances.” (p.724)

With regard to this study, noticing can be interpreted as learners’ attention to the provision of metalinguistic comments, awareness of the discrepancy, intake of metalinguistic knowledge in working memory, and maybe their immediate repair following MCF. Understanding, however, involves storing metalinguistic knowledge in long-term memory, the correct application of target structures in spontaneous production in other instances, and the development of procedural knowledge in the long run. In
comparison to other CF techniques and merely communicative interaction without CF, MCF helps learners to notice the disparity, to analyze and comprehend the linguistic rules in meaningful contexts, and to repair errors. Both levels of awareness are promoted. The facilitative role of MCF in promoting learners’ noticing and understanding can explain why MCF is more effective in interlanguage development from a cognitive perspective. Many studies (Lyster, 2004; Mackey, 2006; Sheen, 2010, 2011; Yang & Lyster, 2010) including the current one manifest the positive relationship between explicit CF, noticing, and learners’ improved language performance.

5.2.2 Uptake and Modified Output

In addition to taxonomies of CF types in terms of explicitness, the other widely accepted categorization of CF is based on whether CF provides opportunities for learners’ modified output proposed by Ellis (2006). MCF is explicit as well as output-prompting, whereas CF such as recast is implicit and input-providing. Output-prompting CF differs from input-providing CF in that the former provides only negative evidence to learners and prompts learners to modify their erroneous output on their own in uptake; in contrast, the latter provides positive evidence and leaves no room for learners’ repair of their errors. Ellis asserts “acquisition is better supported when learners have the opportunity to produce modified output containing the target structure” (2012, p.268). Lyster (2004) even suggests that the ideal way to help the learner convert declarative knowledge into procedural knowledge is through prompting self-correction and modified output. In what follows, the efficacy of MCF in relation to modified output is discussed from both
empirical and theoretical perspectives.

The findings of the present study substantiate previous findings (Ellis et al, 2006; Ellis, 2007; Lei, 2008; Sheen, 2011). The results of the current study revealed that communicative instruction with MCF was more effective than communicative instruction without MCF on both linguistic targets. One of the many factors contributing to the efficacy of MCF may be the beneficial role of modified output prompted by MCF in L2 development. A closer examination of in-class recording data indicated that the MCF group had far more modified output displayed as repair than the control group. Occasionally learners ignored the corrective purpose of MCF moves and did not uptake with repair. Nevertheless, most MCF moves were followed by learners’ modified output and repair in the MCF group on both target structures.

Previous empirical studies also reveal that output-prompting CF is likely to lead to more modified output and be more facilitative of language acquisition than input-providing CF (Ammar & Spada, 2006; Ellis et al, 2006; Lyster, 2004; Sheen, 2007, 2010; Yang & Lyster, 2010). In an earlier observational study, Lyster and Ranta (1997) examined ESL learners’ uptake and repair following 6 types of feedback in French immersion classrooms. They found that output-prompting CF such as MCF, direct elicitation, and clarification requests led to larger amounts of modified output. Input-providing CF in the form of recasts generated the least amount of uptake despite teachers’ preference for recasts in class. The finding of a high rate of modified output prompted by MCF was also substantiated by the current study.

Similarly, Lyster and Mori’s (2006) study suggested that CF types as well as
classroom settings could exert influence on the amount of modified output. They conducted a comparative analysis of class interaction in two different instructional settings—French immersion and Japanese immersion classrooms. They investigated the distribution of and relationship between explicit correction, recasts, prompts and learner uptake and repair. The results indicated that learners’ uptake and repair patterns varied in relation to CF type and classroom settings. French immersion classes showed a large proportion of repair following explicit CF—prompt, while the Japanese immersion classes showed a large proportion of repair from explicit CF—recasts.

Interestingly, Nassaji’s research also produced some mixed results (2007). He examined the relationship between two types of CF and learners’ repair. The CF techniques are output-prompting CF in the form of elicitation and input-providing CF in the form of reformulation. Although Nassaji reported a higher frequency of reformulations (recasts) as in Lyster and Ranta’s study, he found similar amounts of modified output following reformulation (34%) and elicitation (31%). The finding of a comparatively low rate of repair generated by output-prompting CF in the form of elicitation contradicts the findings of some previous research, such as Lyster and Ranta’s. Nassaji suggested that how reformulation and elicitation were implemented had an impact on repair. In Nassaji’s study reformulation and elicitation combined with explicit intonational or verbal prompts led to higher rates of repair.

Although MCF in the present study was not implemented with intonational prompts, it still generated a considerable amount of modified output. Three reasons may account for the high rate of modified output in this study. First, in comparison with other
CF techniques, MCF is highly explicit and output-prompting. These distinctive features of MCF call learners’ attention to errors and impel them to modify their erroneous output. Two, some learners’ failures to modify their output resulted in repeated MCF until learners successfully repaired errors. Multiple MCF moves on one erroneous use of the target structures surely increased learners’ noticing and instances of repair. Third, modified output in the present study was calculated despite its source. Modified output from the learner at whom MCF was directed and modified output from other learners were not distinguished. This calculation could augment the total of modified output following MCF.

The operationalization of MCF in this study was output-prompting CF, i.e., only metalinguistic comments were provided but correct forms were withheld. This is in consensus with the operationalization in most CF studies (eg, Ellis et al, 2006; Lyster, 2004; Lyster and Ranta, 1997), but different from a few CF studies (eg. Li, 2014; Sheen, 2011). In Li’s and Sheen’s studies, MCF was utilized as provision of correct forms first and metalinguistic comments followed. It incorporated both positive evidence and negative evidence. A major drawback of this utilization is the decreasing opportunities for learners’ modified output. Obviously MCF in their studies was still salient but it impeded modified output because the correct form was already provided. Learners did not get any push to modify their errors. However, MCF in the current study provided learners opportunities for hypothesis testing and controlled practice. It allowed learners to achieve a higher level of awareness--understanding linguistic rules through restructuring and reformulating their problematic production in the modified output or repair process.
The facilitative role of MCF operationalized as output-prompting also provides support for theoretical claims (Long, 1983, 1996; Swain, 1995). Stemming from Krashen’s Input Hypothesis, Long stresses the importance of negative feedback in his interaction hypothesis. He argues (1983, 1996) that natural input is insufficient for L2 learning, and interactionally modified input such as CF assists L2 development. Learners need negative evidence to detect the difference between their nontarget-like production and target-like production. The negotiation of form and linguistic modification exists in communication and hence promotes L2 learning. Swain (1995) argues similarly in her Output Hypothesis that output is indispensible to successful language acquisition. She suggests that learners fail to achieve grammatical accuracy if they do not have opportunities for pushed output. The findings that learners receiving MCF outperformed learners receiving no MCF indicated the significant role of modified output in L2 development. Theoretically, it is manifest that learners need opportunities for modified output to increase their accuracy in oral production and to acquire L2 eventually.

Empirically, the present study supports the theoretical claims.

Modified output following MCF is pivotal for EFL learners’ language development. Unlike ESL learners, EFL learners rely largely on classroom instruction to learn the target language due to their limited exposure to natural output outside of the classroom. Researchers acknowledge the significance of systematic and controlled practice in the form of modified output in knowledge development and skill building (DeKeyser, 2007b; Lyster & Izquierdo, 2009; Ranta & Lyster, 2007). Lyster et al. confirm it in this way:
In school-based learning, students need repeated opportunities to retrieve and restructure their knowledge of the target language. This is because students may have target language knowledge that continues to be accessible for comprehension but that requires further activation before becoming readily available for accurate production. CF can be used to revisit target items and grammatical subsystems in ways that encourage the gradual development of a network of associations that become increasingly accessible for learners during communicative interaction. (2013, p.13)

Although modified output and uptake with repair is facilitative of language development, researchers caution that learners’ uptake or modified output cannot be taken as acquisition (Ellis, Basturkmen, & Loewen, 2001; Ellis & Sheen, 2006; Lyster et al, 2013; Sheen, 2006, 2008). Therefore, modified output in the form of repair in this study is only analyzed as a plausible account for the efficacy of MCF rather than an indicator of language acquisition.

5.2.3 Linguistic Targets

The finding that MCF had an overall larger effect on the third person –s than on gender referents indicated that effects of CF were mediated by target structures. MCF exerted a large effect on the third person –s in the elicited oral imitation test over the posttest and the long delayed posttest. MCF exerted medium effects to a variety of degrees on the third person –s in the metalinguistic knowledge test and on gender
pronouns in two test formats and two testing periods. To be specific, the effect of MCF on procedural knowledge is larger than on declarative knowledge on gender referents.

The finding that the efficacy of CF varied in relation to linguistic targets is in line with previous research (Ellis, 2007; Li, 2014; Long, Inagaki and Ortega, 1998; Yang & Lyster, 2010). For example, Ellis (2007) compared the differential effects of recasts and MCF on the acquisition of two grammatical structures -- comparative -er and past tense -ed in a quasi-experimental study. Statistical results indicated that MCF had a greater effect on comparative -er than on past tense -ed; however, recasts did not promote the acquisition of the two target structures.

Additionally, Yang and Lyster (2010) provided support for the differential effects of CF on two grammatical structures. They compared the effects of recasts and prompts including MCF on Chinese college EFL learners’ acquisition of regular and irregular past tense. They found that prompts were more effective than recasts on the use of regular past tense forms; whereas prompts and recasts did not exhibit difference on irregular past tense forms.

Li’s (2014) most recent study on the interface between CF, learner proficiency, and two Chinese linguistic targets rendered some subtle findings. He chose two very different target structures – Chinese perfective -le and classifiers. The results showed that 1. MCF was effective for both structures and for learners at both low and high proficiency levels; 2. MCF was more effective on both linguistic targets for low-level learners than recasts; 3. recast was effective for the low-level learners’ acquisition of classifiers but not perfective -le; 4. the effect of recasts was enhanced over time for
perfective -le at the high proficiency level, but not for classifiers.

In the present study, MCF was effective on both linguistic targets despite test formats. A closer examination of effect sizes revealed that MCF was more effective on learning third person –s than learning gender referents. This finding suggested the nature of linguistic structures could mediate the efficacy of MCF. To account for this, an examination of the nature of target structures is necessary.

According to Ferris and Roberts (2001), gender referent errors fall in the category of lexical error. Errors in word choice or word form, for example pronoun and preposition errors, all fit into this category. In terms of hindrance to communication, gender referent errors are global errors because they cause communication breakdowns and interfere with understanding (Burt, 1975). Spoken Chinese has only one word ta referring to he, she, him, and her. Gender message in gender referents in spoken Chinese is rather blurred. George (1972) believes when learners’ L1 has one word for third person singular pronouns, they follow the L1 model and are likely to use either he or she for both he and she unaware of a reduction in the gender information of English sentences. In contrast, third person –s errors are form-based or rule-based grammatical errors. They are local errors because they do not cause serious communication breakdowns. Third person –s structure ranks at the top on the acquisition/difficulty order (Dulay & Burt, 1973, 1974; Stockwell, Bowen & Martin, 1965; Zobl & Liceras, 1994). Morpheme –s appears redundant because it does not convey meaning despite its morphosyntactic function. Ammar (2008) states that “a form is said to be redundant when it does not contribute to the meaning being communicated because the meaning of that same form is already
conveyed by another element of the sentence” (p.202). Other researchers such as DeKeyser (2005) and Spada and Tomita (2010) also underpin this viewpoint. They hold that the weak form-meaning mapping is a source of difficulty for third person –s.

Based on the above analysis, it is clear that learners’ errors in gender referents are deduction of meaning and replacement of other case forms whereas third person –s errors are merely deduction of form or replacement of wrong form. Meaning is not a concern in third person –s errors. Although third person –s ranks top on the acquisition/difficulty hierarchy, the rule for third person –s is simple and straightforward. Therefore, the cognitive burden in processing gender referents is more complex than third person –s. The reason is that the former requires attention to both meaning and form while the latter only requires attention to form. MCF learners making gender referent errors may be able to reach the first level of “noticing” and repair subsequently upon receiving MCF. However, it may be more difficult for them to reach the second level --“understanding”, i.e., correctly applying the meta-language to other instances. This may explain the larger effect on third person –s than on gender referents given that learners did the same communicative tasks for both target structures.

In addition, this finding substantiated as well as contradicted some findings in Mackey and Goo’s (2007) meta-analysis study. After pooling 16 studies that investigated the effect of interactional feedback on L2 learning, Mackey and Goo found that interactional feedback was more beneficial for lexis than for grammar in the short-term but more facilitative for grammar in the long-term. Notwithstanding, the current study evidenced an overall bigger effect size for third person -s than for gender pronouns.
regardless of time. One possible explanation could be the different categorization of grammar and lexis. As stated earlier, this study follows Ferris’ classification of pronouns as lexis. In Lyster’s (1998) study, however, pronoun errors are classified as grammar errors. He considered “errors in closed classes as grammatical and errors in open classes as lexical” (p.195). Nevertheless, it is obvious that third person -s differs from gender pronouns in terms of their linguistic features and the effect of MCF varies in relation to linguistic targets.

5.3 Declarative Knowledge and Procedural Knowledge

Learners’ accuracy scores in the elicited imitation test and the metalinguistic test suggested that learners’ procedural knowledge on both linguistic structures improved significantly. Learners’ declarative knowledge on both structures also improved, but not as dramatically as procedural knowledge. The result indicates that MCF facilitates the proceduralization of declarative knowledge, i.e., MCF plays a positive role in bridging the gap between declarative knowledge and procedural knowledge on both structures. This finding is in accordance with previous theoretical and empirical claims (DeKeyser, 1998, 2007a,b; Ellis et al., 2006; Lyster, 2004).

Skill acquisition theory in SLA holds that language learning resembles other human learning process in which learners go through three stages to acquire L2 -- declarative, procedural, and autonomous stages. DeKeyser (2007b) explains “the learning of a wide variety of skills shows a remarkable similarity in development from initial representation of knowledge through initial changes in
behavior to eventual fluent, spontaneous, largely effortless, and highly skilled
behavior, and that this set of phenomena can be accounted for by a set of basic
principles common to acquisition of all skills” (p. 97).

Many researchers agree that repeated and sufficient practice generated from CF can
gradually convert declarative knowledge to procedural knowledge, i.e., proceduralization
(DeKeyser, 1998, 2007a; Ellis et al., 2006; Lyster & Sato, 2012). Learners in the current
study exhibited good command of declarative knowledge of the target structures, but less
satisfactory command of procedural knowledge in the pretest. Through MCF embedded
in meaningful communicative tasks, their procedural knowledge improved significantly
over time. In other words, they were moving from declarative stage to procedural stage
with the help of MCF in developing their communicative skills of the target structures.

Another theoretical claim relevant to declarative knowledge and procedural
knowledge is the interface hypothesis. It concerns the perception of how declarative
knowledge interfaces with procedural knowledge. Of the three positions, the non-
interface position posits that declarative knowledge and procedural knowledge are
processed through two independent mechanisms (Krashen, 1981; Hulstijn, 2002).
Declarative knowledge cannot be transformed into procedural knowledge. The strong
interface position contends that declarative knowledge can be developed into procedural
knowledge through plentiful communicative practice (DeKeyser, 1998, 2007a; Sharwood
Smith, 1981). The weak interface position proposes that declarative knowledge and
procedural knowledge can work cooperatively and are convertible under some
conditions, for example learners readiness, noticing linguistic structure and noticing the
gap between the input and their interlanguage. The current research result provides support for the interface position and Bialystok’s (1978) claim that declarative knowledge and procedural knowledge are on a continuum. The two knowledge types can be converted under certain conditions. The research result suggests that in addition to those conditions mentioned above, repeated practice and comparatively long time are also indispensable. The long-term research design is equally if not more important to these Chinese EFL learners’ time-consuming proceduralization of the target structures. The provision of MCF over an extended period of time made it possible for declarative knowledge to interface with procedural knowledge.

Through communicative form-focused tasks in combination with MCF, the experimental group in this study got more opportunities to modify their erroneous output and practice the target structures than the control group. Systematic, consistent, and focused MCF on target structures over an extended period of time provided EFL learners with repeated practice opportunities, which would otherwise be missing in regular EFL instructions. Consequently, MCF served as a feasible and effective means to help learners gradually convert their declarative knowledge to procedural knowledge.

In addition to the theoretical support, this study together with other empirical studies also lends support for the facilitative role of CF in the development of procedural knowledge. DeKeyser (1997) reported the conducive role of practice in improving learners’ accuracy and automatization in use of grammar rule in artificial language. Lyster’s (2004) classroom-based study found that F-on-F combined with output-promoting CF was more effective than F-on-F combined with input-providing CF or no
CF in proceduralizing immersion L2 learners’ knowledge of French grammatical gender.

Likewise, this finding is supported by the study conducted by Ellis et al. (2006). They proved the beneficial role of MCF in developing both declarative knowledge and procedural knowledge of English past tense –ed. He explains the role of MCF this way, “time-outs from communicating afforded by explicit correction constitute a perfect context for melding the conscious and unconscious processes involved in learning” (p. 343). In other words, MCF allows both declarative stage and procedural stage to be activated and employed in the learning process.

In the current study, learners improved their declarative knowledge together with their procedural knowledge, however, to a less extent. This may be attributed to the ceiling effect of learners’ accuracy scores in the metalinguistic test. All these learners attended middle schools in China. These two linguistic targets had been taught explicitly before they entered college. It is highly likely that these intermediate Chinese college EFL learners had developed declarative knowledge of the target structures but had not proceduralized it yet. Evidence is learners’ high outset score in the metalinguistic knowledge test and low outset score in the elicited oral imitation test in the pretest period. As a result, it was hard to see learners’ dramatic improvement in later testing periods.

The research finding suggests that MCF can bridge the gap between declarative knowledge and procedural knowledge of gender referents and third person –s. Moreover, it provides additional support for researchers’ call for measuring both declarative and procedural knowledge in CF research (Ellis, 2009; Sheen, 2011). Obviously, it is of great importance to tease out the two types of knowledge in evaluating the effectiveness of CF.
5. 4 Pedagogical Implications

The current study found that learners who received MCF embedded in communicative tasks improved their procedural knowledge more dramatically than the learners only performing communicative tasks without MCF on third person –s. This finding sheds light on the learnability of difficult structures. Third person –s has been widely labeled as one of the most difficult structures for English L2 learners. Some researchers claim that difficult structures become unacquirable for adult learners because of the ending of a critical period for language learning (Hawkins & Chan, 1997; Lardiere, 1998a, b). According to them, the difficult structures are fossilized for adult learners. Nevertheless, this study produced a more promising finding. Learners’ procedural knowledge on third person –s was found to improve dramatically over time regardless of their moderate improvement in declarative knowledge. This suggests that language teachers should not give up on students’ errors no matter how persistent or serious they are. Errors need to be attended to and treated in language classrooms despite learners’ exposure to positive evidence elsewhere.

Although language practitioners are found to be in favor of implicit CF such as recasts in classrooms, a great many studies exploring the effect of CF on L2 development demonstrate that explicit/output-prompting CF such as MCF is more beneficial. The result of the present study is congruent with the many studies that provide statistical support for the efficacy of MCF in language classrooms. MCF embedded in communicative tasks was found to promote learners’ noticing and prompt modified output in comparison to communicative tasks alone. From the perspective of learners, the
data from the exit questionnaire clearly reveal learners’ preference for MCF. The very few hesitant and ambiguous voices from learners even acknowledged the benefits of learning from CF. Their chief concern was rather affective accounts, as indicated in the questionnaire that they felt embarrassed or shy to be corrected publicly in front of the class. Therefore, language teachers are advised to make use of MCF in classrooms for learners’ benefits. When teachers choose CF techniques, it is advisable that they make use of explicit and output-prompting CF in order to promote learners’ noticing and enlarge learners’ opportunities of modified output. This is particularly helpful to EFL learners who have limited exposure and access to an authentic English environment.

Additionally, teachers need to take the nature of linguistic structures into consideration when selecting CF techniques given that MCF exerted differential effect sizes on different linguistic structures in this study.

The pretest score of the research participants showed that they did not have a balanced declarative knowledge and procedural knowledge of both linguistic targets. Their procedural knowledge of the two structures lagged far behind their declarative knowledge. Such imbalance was also evidenced in learners’ speaking proficiency. De Jong et al. (2012) found from their study that speaking proficiency was determined by linguistic knowledge, linguistic processing knowledge, and pronunciation skills. According to them, the first two components are synonymous with declarative knowledge and procedural knowledge. Evidently, both declarative knowledge and procedural knowledge are indispensible constructs in L2 proficiency. Metalanguage is generally associated with the knowledge of *what*, i.e., declarative knowledge. The knowledge of
how constitutes procedural knowledge. It is thus postulated that metalanguage can only improve declarative knowledge but not procedural knowledge or L2 proficiency in general. Notwithstanding, this study demonstrated that metalanguage conveyed through MCF played a more facilitative role in promoting intermediate-level learners’ procedural knowledge than declarative knowledge.

Language teachers can interpret it as the necessity to have and employ metalanguage in class even with learners who already have good mastery of declarative knowledge. In other words, good command of declarative knowledge does not necessarily qualify learners as proficient speakers, while procedural knowledge is essential in speaking proficiency and declarative knowledge may evolve into procedural knowledge with the help of metalanguage. Teachers can design a class syllabus that tactfully involves metalanguage in communicative contexts in order to promote intermediate-level learners’ L2 proficiency. However, language teachers need to be careful when they apply this research finding to learners at other levels. Metalanguage was proved beneficial to learners who have learned the target structures and have good declarative knowledge in this study. It remains unclear whether metalanguage is still beneficial to learners who do not have declarative knowledge yet.

To conclude this section, MCF is necessary but not sufficient to bridge the gap between declarative knowledge and procedural knowledge. Ample opportunities for practice of target structures are pivotal to the interface of language learners’ declarative and procedural knowledge.
5. 5 Limitations

Despite the best hope for a well-controlled study, the current study has several limitations, including the convenience sampling, the treatment length, and the testing period. These limitations are commonplace in a study of this nature. However, none of them adversely impacted the results of the current study.

First, this study used convenience sampling. The research participants came from intact classes in the same university. The university admitted the students based on their scores in national college entrance examination. Therefore, the generalizability of the study to other population needs to be considered carefully. The research findings observed on these intermediate-level college EFL learners may not be replicated on other English learners. The involvement of learners from other universities may bring about a more holistic picture.

Second, the treatment length may not be sufficient. The study was a classroom-based quasi-experimental study. The treatment was designed following the course syllabus, embedded and completed in regular class instruction. This may make instructional treatments and research findings replicable and applicable to other similar pedagogical settings in comparison to a lab-based study. However, the real treatment sessions did not last long because of the commitment to other class agendas on the course syllabus. The total treatment length is 689 minutes, 352 minutes for the MCF group and 337 minutes for the control group. According to some meta-analysis studies of CF, length of treatment over 120 minutes is regarded as long in Li’s study (2010) and length over 7 hours is regarded as long in Norris and Ortega’s study (2000) as well as Lyster and
Saito’s study (2010). The experimental group barely reached the bottom-line according to two meta-analysis studies. However, it should be acknowledged that the proceduralization of declarative knowledge takes a long time and hence longer treatment will be desirable.

Third, incidental CF on linguistic targets was not considered in this study but it may also have an impact on learners’ interlanguage development. The teacher incidentally provided MCF on errors in use of linguistic targets emerging in other class activities. Because CF in this case was random and rare, these CF moves were excluded from the data analysis.

Fourth, the long interval between the end of treatment and the long delayed posttest may leave room for other interventions to influence learners’ performance in the delayed posttest. The long delayed posttest design was intended to respond to Norris and Ortega’s (2000) and Truscott’s (1998) suggestion for long delayed posttests in CF research. The long interval may allow for a valid examination of the retention of the effect of CF, but it may as well pose a validity threat of an intervening variable between the end of the instructional treatment and the long delayed posttest. Shortening the interval between the end of instructional treatment and the delayed posttest may solve the problem, but it may not be able give a valid examination of the effect retention. Effects may not fade immediately but over an extended period of time.

In summary, researchers and teachers need to consider these limitations when interpreting the research findings.
5.6 Suggestions for Future Research

As Ellis states “the extent to which the teacher’s use of metalanguage assists language acquisition remains largely unstudied to date” (2012, p. 134), The current knowledge about the role MCF plays in developing learners’ declarative knowledge and procedural knowledge is still limited. There is a lot remaining unclear and the topic is worth exploring further.

First, studies investigating proceduralization in relation to other linguistic structures could be conducted to increase knowledge about proceduralization of declarative knowledge. As known, the nature of linguistic structures can pose different levels of learning difficulty and greatly impact success of language acquisition. More knowledge of how and which linguistic structures interrelate with proceduralization is needed.

Second, proceduralization in relation to learners’ individual factors such as readiness and age could be investigated. Questions such as to what degree declarative knowledge is sufficient for future proceduralization, how does declarative knowledge influence the development of procedural knowledge, does age have an impact on converting declarative knowledge to procedural knowledge, will be interesting topics in future research.

Third, the study of the effectiveness of MCF in comparison with other CF techniques in other instructional settings may enrich the current limited knowledge and contribute to determining generalizability and applicability to other class contexts.

Fourth, more exploration of other feasible instruction that may help promote the
development of procedural knowledge is needed. So far, the knowledge about how procedural knowledge can be promoted in L2 classrooms is limited. There are still a lot worth exploring.

### 5.7 Final Conclusions

The present study adds to limited knowledge of the role of Metalinguistic Corrective Feedback in developing EFL learners’ declarative knowledge and procedural knowledge in three aspects. First, MCF embedded within communicative form-focused tasks exerted a positive effect on developing learners’ declarative knowledge and procedural knowledge. Moreover, the effect was retained over an extended period of time. Second, MCF exerted a larger effect on developing procedural knowledge than declarative knowledge regardless of linguistic targets. Third, MCF was more effective in improving learners’ accuracy in use of the morphosyntactic structure third person –s than in use of the meaning-based structure gender referents.

It may be too rushed to draw the conclusion that the gap between declarative knowledge and procedural knowledge can be closed through MCF simply based on this study. The findings obtained from the current study do provide support for the efficacy of MCF in proceduralizing both the meaning-based structure gender referents and the rule-based structure third person morpheme –s for intermediate Chinese EFL learners. Sheen states in her book “the benefits of metalinguistic feedback may have been underestimated, especially in contexts where Communicative Language Teaching and implicit knowledge/learning are prioritized” (2011, p.74). In general, the study confirmed
the positive role of MCF in proceduralization of gender referents and third person –s in a communicative classroom in a Chinese university. Therefore, EFL teachers could consider the use of MCF for their students’ benefit.
Appendix 1 Informed Consent Form

INFORMED CONSENT FORM

I am a doctoral student from School of Education at Boston University, USA. I will conduct my research in your class to complete my doctoral requirements. I have been granted permission to conduct the research at Dalian University of Foreign Languages by the school deans, Dr. Junyue Chang and Dr. Fengguang Liu. In addition, I need your consent.

The purpose of this study is to examine the feasible strategy to bridge declarative knowledge and procedural knowledge of two English linguistic features. Insights will be used in writing a dissertation, which will be read by my dissertation committee at Boston University’s School of Education.

Your class discussion will be audiotaped with digital tape recorders to help me accurately capture your use of some linguistic structures. The recording will be listened and transcribed by me. However, the recording and the transcription will remain anonymous and will not be identified to individuals. The only purpose of the recording is for me to count the whole class’s errors and the error-correction of the targeted English linguistic items. You will take knowledge tests of the target features for about 30 minutes. The tests will be anonymous and none of the tests will be counted as a part of your course grade.

This study to learn the feasible strategy to reduce college students’ speech errors in English has no inherent risk. However, you have the right to withdraw from the study at any time. In the event you choose to withdraw from this study, I will destroy all the information you provide and omit from the final written dissertation.

Please feel free to ask questions or raise concerns at any time. You may contact me, Qin Wang, at melodyqw@bu.edu or 1333 222 5701. You may also contact the Dean, Fengguang Liu, at emilylfg@yahoo.com.cn or 86 111 021. You may obtain further information about your rights as a research participant by calling the BU CRC IRB Office at 1-617-358-6115.

Thank you in advance for participating in this study!

By signing this consent form, I certify that I ________________________ (Print full name here) agree to the terms of this agreement.

__________________________  _______________________
(signature)                    (date)
Appendix 2 Pretest--Elicited Oral Imitation Test

Elicited Oral Imitation Test

Part I

1. The violin is one of the most difficult musical instruments to learn.
2. A good hostess makes friends at home when he invites them home.
3. A waitress gets more tips if he looks good.
4. Beethoven composed many works even when he turned deaf.
5. Animals would enjoy a better life without human beings.
6. A woman cares more about her children than his husband.
7. Princess Diana still loved her husband after divorcing her.
8. Jackie Chen started her career in Hollywood when he was 40.
9. Snakes only hibernate in winter but not in summer.
10. An actor has to be handsome; otherwise he won’t be popular.

Part II

11. An Apple Macbook cost much more than a Dell netbook does.
12. Rich people should donate more money to poor people.
13. Not every criminal regret his wrongdoing and feels sorry when caught.
14. Most diplomats have college degrees in political science.
15. Our government values education and everyone knows that.
16. A salesperson always make good suggestions to customers.
17. Yang Zhenning’s second marriage shocked the whole country.
18. A nation’s culture pass down from one generation to another.
19. A good coach make training fun and cares about players.
20. It is important to teach our children Chinese history.
Appendix 3 Pretest—Metalinguistic Knowledge Test

Metalinguistic Knowledge Test

(Take the time you need to complete the test.)

Part I
Directions: Please read the following sentences and select the best choice among the four.

1. Mary is such a nice person. He drives me back on her way home every day.
Which best describes the error in the sentences? _________.
   A. Tense error       B. Gender referent error
   C. Preposition error D. Wrong choice of  word

2. Here peaches cost one dollar per pound and milk cost three dollars a carton.
Which best describes the error in the sentence? _________.
   A. subject error     B. noun error
   C. adverb error      D. third person –s error in simple present tense

3. A person who always _______ his past needs to look forward.
   A. recalls           B. recall
   C. recalling         D. have recalled

4. George is a good father. His children love ______ very much.
   A. her       B. she       C. it       D. him

5. Which of the following is a gender referent?
   A. she       B. book      C. run      D. funny

6. Which of the following underlined part is third person –s in simple present tense?
   A. She stops by her parents’ house every day.
   B. Murphy is the best student in the class.
   C. We have many toys in our house.
   D. The player was so angry.
Part II

7. Underline gender referents in the following sentences.

She told us that this Sunday would be a really important day in her husband’s life. He and his team worked so hard to get to this point and now they need friends more than ever to send them positive energy so they can fulfill their dream. She has supported her husband ever since he made the decision to get it done.


Mr. Lee is a bus driver. Every day he gets up at 7:00 a.m. But today he got up late. Usually he showers, eats his breakfast, and drives to the station. But today he didn’t take a shower and his wife drove him to work. He often drinks coffee and listens to the radio on his way to work. But today he didn’t drink anything. He calls his wife after work every day. But today he didn’t.
Appendix 4 Posttest—Elicited Oral Imitation Test

Elicited Oral Imitation Test

Part I

1. People do not have to go to school to receive an education.
2. Madame Curie is a great scientist and she won many awards.
3. Generally a man likes sports much more than her girlfriend.
4. Students who have part-time jobs have poorer grades than those who don’t.
5. A man cares less about his children than her wife.
6. Yao Chen became more famous after divorcing her husband.
7. Lu Xun changed his major to literature when she was 20.
8. If seawater is drinkable, the whole world will be happier.
9. A girl prefers shopping with her girlfriend not with her boyfriend.
10. An actress has to be pretty; otherwise he won’t be popular.

Part II

11. People should report their stolen money and phone to the police.
12. A surgeon works longer hours and earns more money than a dentist.
13. Everyone loves comic books and read them.
14. A good teacher make lessons interesting and cares about students.
15. Michael Jackson is a great singer and many people like him.
16. A good education benefit people and the society people live in.
17. The construction of our new campus was completed last year.
18. A young man like cigarettes and fast cars more than a young woman do.
19. The resources on the earth will run out in the future.
20. To most people, a happy day makes an efficient day.
Appendix 5 Posttest—Metalinguistic Knowledge Test

Metalinguistic Knowledge Test

(Take the time you need to complete the test.)

Part I
Directions: Please read the following sentences and select the best choice among the four.

1. Lily is a wonderful friend. He remembers all her friends’ birthday.
   Which best describes the error in the sentences? __________.
   A. Tense error   B. Gender referent error
   B. C. Preposition error   D. Wrong choice of word

2. Driving from home to school only take me hour an hour.
   Which best describes the error in the sentence? __________.
   A. subject error   B. noun error   C. adverb error
   D. third person –s error in simple present tense

3. A person who always _______ may bring negative influence on others.
   A. complains   B. complain   C. complaining   D. have complained

4. Diana is a good mother. Her children love ______ very much.
   A. her   B. it   C. his   D. him

5. Which of the following is a gender referent?
   A. him   B. calendar   C. escape   D. interesting

6. Which of the following underlined part is third person –s in simple present tense?
   A. He visits his parents every two months.   B. Picasso is one of the best artists.
   C. They have many books in the house.   D. The scientist was so pessimistic.
Part II

7. Underline gender referents in the following sentences.

Bill and Melinda Gates met in 1987 at a Microsoft press event in Manhattan. Bill proposed marriage to her in 1993. They have three children. Planning on leaving his children $10 million each, Gates has always said that, like Carnegie, he will give away most of his fortune before he dies. Gates plans to make sure his children are well taken care of but doesn’t want to leave them the burden of tremendous wealth. Melinda totally agrees with her husband’s idea.


When my mum goes shopping, she always buys rice and vegetables. Sometimes my little sister goes with her. My little sister likes chocolate, and I like ice cream, so mum also gets these things for us. When she comes home, we always help her in the kitchen. I wash the vegetables and mum cooks the dinner.
Appendix 6 Short Exit Questionnaire

Short Exit Questionnaire

1. Now that you have completed the tests, what do you think they were all about?
   a) They were practicing and testing my writing.
   b) They were practicing and testing my grammar.
   c) They were practicing and testing my listening and speaking.
   d) They were practicing and testing my vocabulary.

2. Do you like to be corrected while you are speaking in class? Why or why not?

3. Please write a couple of sentences saying what you think you learned from the class activities and the tests.
Appendix 7 Delayed Posttest—Elicited Oral Imitation Test

Elicited Oral Imitation Test

Part I
1. A mom values his children more than her husband.
2. The piano is one of the most difficult musical instruments to learn.
3. A man likes to go shopping much more than her girlfriend.
4. Hillary Clinton is a successful leader and people love her.
5. Blue collars earn less than white collars in most countries.
6. George Bush stayed in the White House longer than her dad.
7. Mozart was famous in that he’s a talented musician.
8. Yao Ming started his basketball career at eight.
9. Though Steve Jobs passed away, she’s remembered worldwide.
10. A woman cares more about her family than his husband.

Part II
11. Everybody know water freezes at zero degrees.
12. Rich people should donate more money to poor people.
13. Lang Lang enjoy playing the piano now, but not before.
14. A good English speaker gets more job opportunities.
15. Shanghai is the biggest city and lies in the east of China.
16. An apple a day provides us with enough vitamin C.
17. People can get education without going to school.
18. The sun rise in the east and set in the west.
19. It rains less in spring than in summer in Dalian.
20. Han Han live in Beijing for easy access to publishers.
Appendix 8 Delayed Posttest -- Metalinguistic Knowledge Test

Metalinguistic Knowledge Test
(Take the time you need to complete the test.)

Part I
Directions: Please read the following sentences and select the best choice among the four.

1. Rose is a wonderful friend. He remembers all of his friends’ birthday. Which best describes the error in the sentences? ________.
   A. Tense error       B. Gender referent error
   C. Preposition error D. Wrong choice of word

2. It takes about an hour to fly from Dalian to Beijing. Which best describes the error in the sentence? ________.
   A. subject error     B. noun error
   C. adverb error      D. third person –s error in simple present tense

3. A negative teacher ______ negative influence on students.
   A. brings  B. bring  C. bringing  D. have brought

4. John is an amazing priest. All the church people love ______ very much.
   A. her        B. she     C. they    D. him

5. Which of the following is a gender referent?
   A. her    B. book    C. fly    D. wonderful

6. Which of the following underlined part is third person –s in simple present tense?
   A. She goes to the library every two weeks.
   B. Picasso is one of the best artists.
   C. There are many CDs on the shelf.
   D. The scientist gave a wonderful speech.
Part II

7. Underline gender referents in the following sentences.
Mr. French, now 71, has a little-known and frequently misdiagnosed group of brain diseases that eat away at personality and language. Looking back, Mrs. French recalled episodes of odd behavior over the years and realized that her husband’s mind had probably begun to slip while he was in his 50s. He had always changed jobs a lot. At the time she took it as a sign of a stubborn personality, not of illness — and it is still not clear which it was. She thought of divorce but changed her mind after they got the diagnosis.

An old tiger lives in the forest. He often tells other animals to bring him something to eat. He sees a monkey and says, “I’m hungry. Go to the village and bring me a fat pig.” “Oh, Tiger,” says the monkey. “I can’t do that now. There is another tiger over there. He also wants a fat pig. He will not let me get anything for you to eat. I am afraid of him.” “What?” cries the tiger. “Show me that tiger. I’ll talk to him.”
Appendix 9 Student Copy of Elicited Oral Imitation Test (pretest, posttest, and delayed posttest)

Directions: You will hear altogether 20 sentences. Please listen carefully and pay attention to the meaning of each sentence. Then you will have five seconds to mark down your opinions about the statements on your answer sheet Agree, Disagree, or Do Not Know. After you hear a beep, repeat verbatim what you hear previously in correct English.

NOTE: Some sentences you hear may not be read in correct English. Please make corrections as needed. You are not allowed to take notes in the test.

Part I
1. Agree       Disagree       Do not know
2. Agree       Disagree       Do not know
3. Agree       Disagree       Do not know
4. Agree       Disagree       Do not know
5. Agree       Disagree       Do not know
6. Agree       Disagree       Do not know
7. Agree       Disagree       Do not know
8. Agree       Disagree       Do not know
9. Agree       Disagree       Do not know
10. Agree      Disagree       Do not know

Part II
11. Agree      Disagree       Do not know
12. Agree      Disagree       Do not know
13. Agree      Disagree       Do not know
14. Agree      Disagree       Do not know
15. Agree      Disagree       Do not know
16. Agree      Disagree       Do not know
17. Agree      Disagree       Do not know
18. Agree      Disagree       Do not know
19. Agree      Disagree       Do not know
20. Agree      Disagree       Do not know
### Appendix 10 Statistic Results Through SPSS 22

Table 10.1 ANOVA result for gender referents in the elicited oral imitation test

<table>
<thead>
<tr>
<th></th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>gender pre</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Between Groups</td>
<td>28.967</td>
<td>1</td>
<td>28.967</td>
<td>.195</td>
<td>.661</td>
</tr>
<tr>
<td>Within Groups</td>
<td>8629.366</td>
<td>58</td>
<td>148.782</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>8658.333</td>
<td>59</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>gender post</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Between Groups</td>
<td>1683.739</td>
<td>1</td>
<td>1683.739</td>
<td>7.110</td>
<td>.010</td>
</tr>
<tr>
<td>Within Groups</td>
<td>13734.594</td>
<td>58</td>
<td>236.803</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>15418.333</td>
<td>59</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>gender dpt</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Between Groups</td>
<td>1690.814</td>
<td>1</td>
<td>1690.814</td>
<td>4.633</td>
<td>.036</td>
</tr>
<tr>
<td>Within Groups</td>
<td>21167.519</td>
<td>58</td>
<td>364.957</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>22858.333</td>
<td>59</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 10.2 ANOVA result for gender referents in the metalinguistic knowledge test

<table>
<thead>
<tr>
<th></th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>gen</td>
<td>Between Groups</td>
<td>77.152</td>
<td>1</td>
<td>77.152</td>
<td>.449</td>
</tr>
<tr>
<td>der</td>
<td>Within Groups</td>
<td>9962.848</td>
<td>58</td>
<td>171.773</td>
<td></td>
</tr>
<tr>
<td>pre</td>
<td>Total</td>
<td>10040.000</td>
<td>59</td>
<td></td>
<td></td>
</tr>
<tr>
<td>gen</td>
<td>Between Groups</td>
<td>600.224</td>
<td>1</td>
<td>600.224</td>
<td>2.699</td>
</tr>
<tr>
<td>der</td>
<td>Within Groups</td>
<td>12898.109</td>
<td>58</td>
<td>222.381</td>
<td></td>
</tr>
<tr>
<td>post</td>
<td>Total</td>
<td>13498.333</td>
<td>59</td>
<td></td>
<td></td>
</tr>
<tr>
<td>gen</td>
<td>Between Groups</td>
<td>213.053</td>
<td>1</td>
<td>213.053</td>
<td>1.636</td>
</tr>
<tr>
<td>der</td>
<td>Within Groups</td>
<td>7551.947</td>
<td>58</td>
<td>130.206</td>
<td></td>
</tr>
<tr>
<td>dpt</td>
<td>Total</td>
<td>7765.000</td>
<td>59</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 10.3 Repeated measures ANOVA result for gender referents in the elicited oral imitation test

<table>
<thead>
<tr>
<th>Source</th>
<th>Sphericity</th>
<th>Type III Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>time</td>
<td>Sphericity Assumed</td>
<td>4025.604</td>
<td>2</td>
<td>2012.802</td>
<td>8.859</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td>Greenhouse-Geisser</td>
<td>4025.604</td>
<td>1.927</td>
<td>2088.748</td>
<td>8.859</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td>Huynh-Feldt</td>
<td>4025.604</td>
<td>2.000</td>
<td>2012.802</td>
<td>8.859</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td>Lower-bound</td>
<td>4025.604</td>
<td>1.000</td>
<td>4025.604</td>
<td>8.859</td>
<td>.004</td>
</tr>
<tr>
<td>time * group</td>
<td>Sphericity Assumed</td>
<td>1438.937</td>
<td>2</td>
<td>719.469</td>
<td>3.167</td>
<td>.046</td>
</tr>
<tr>
<td></td>
<td>Greenhouse-Geisser</td>
<td>1438.937</td>
<td>1.927</td>
<td>746.615</td>
<td>3.167</td>
<td>.048</td>
</tr>
<tr>
<td></td>
<td>Huynh-Feldt</td>
<td>1438.937</td>
<td>2.000</td>
<td>719.469</td>
<td>3.167</td>
<td>.046</td>
</tr>
<tr>
<td></td>
<td>Lower-bound</td>
<td>1438.937</td>
<td>1.000</td>
<td>1438.937</td>
<td>3.167</td>
<td>.080</td>
</tr>
<tr>
<td>Error (time)</td>
<td>Sphericity Assumed</td>
<td>26356.618</td>
<td>116</td>
<td>227.212</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Greenhouse-Geisser</td>
<td>26356.618</td>
<td>111.782</td>
<td>235.785</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Huynh-Feldt</td>
<td>26356.618</td>
<td>116.000</td>
<td>227.212</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Lower-bound</td>
<td>26356.618</td>
<td>58.000</td>
<td>454.424</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Tests of Between-Subjects Effects

<table>
<thead>
<tr>
<th>Source</th>
<th>Type III Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>221788.194</td>
<td>1</td>
<td>221788.194</td>
<td>2246.955</td>
<td>.000</td>
</tr>
<tr>
<td>group</td>
<td>654.861</td>
<td>1</td>
<td>654.861</td>
<td>6.634</td>
<td>.013</td>
</tr>
<tr>
<td>Error</td>
<td>5724.954</td>
<td>58</td>
<td>98.706</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 10.4 Pairwise comparison of the accuracy scores on gender referents in the elicited oral test

**Tests of Within-Subjects Contrasts**

<table>
<thead>
<tr>
<th>Source</th>
<th>time</th>
<th>Type III Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Level 1 vs. Level 2</td>
<td>5834.401</td>
<td>1</td>
<td>5834.401</td>
<td>15.638</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td>Level 2 vs. Level 3</td>
<td>6.674</td>
<td>1</td>
<td>6.674</td>
<td>.014</td>
<td>.905</td>
</tr>
<tr>
<td>time * group</td>
<td>Level 1 vs. Level 2</td>
<td>2154.401</td>
<td>1</td>
<td>2154.401</td>
<td>5.775</td>
<td>.019</td>
</tr>
<tr>
<td></td>
<td>Level 2 vs. Level 3</td>
<td>.007</td>
<td>1</td>
<td>.007</td>
<td>.000</td>
<td>.997</td>
</tr>
<tr>
<td>Error(time)</td>
<td>Level 1 vs. Level 2</td>
<td>21638.932</td>
<td>58</td>
<td>373.085</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Level 2 vs. Level 3</td>
<td>26993.326</td>
<td>58</td>
<td>465.402</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 10.5 Repeated measures ANOVA result for gender referents in the metalinguistic knowledge test

<table>
<thead>
<tr>
<th>Source</th>
<th>Type III Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>time</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sphericity Assumed</td>
<td>1879.616</td>
<td>2</td>
<td>939.808</td>
<td>5.376</td>
<td>.006</td>
</tr>
<tr>
<td>Greenhouse-Geisser</td>
<td>1879.616</td>
<td>1.813</td>
<td>1036.485</td>
<td>5.376</td>
<td>.008</td>
</tr>
<tr>
<td>Huynh-Feldt</td>
<td>1879.616</td>
<td>1.901</td>
<td>988.783</td>
<td>5.376</td>
<td>.007</td>
</tr>
<tr>
<td>Lower-bound</td>
<td>1879.616</td>
<td>1.000</td>
<td>1879.616</td>
<td>5.376</td>
<td>.024</td>
</tr>
<tr>
<td>time * group</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sphericity Assumed</td>
<td>126.282</td>
<td>2</td>
<td>63.141</td>
<td>.361</td>
<td>.698</td>
</tr>
<tr>
<td>Greenhouse-Geisser</td>
<td>126.282</td>
<td>1.813</td>
<td>69.636</td>
<td>.361</td>
<td>.677</td>
</tr>
<tr>
<td>Huynh-Feldt</td>
<td>126.282</td>
<td>1.901</td>
<td>66.432</td>
<td>.361</td>
<td>.687</td>
</tr>
<tr>
<td>Lower-bound</td>
<td>126.282</td>
<td>1.000</td>
<td>126.282</td>
<td>.361</td>
<td>.550</td>
</tr>
<tr>
<td>Error(time)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sphericity Assumed</td>
<td>20279.273</td>
<td>116</td>
<td>174.821</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Greenhouse-Geisser</td>
<td>20279.273</td>
<td>105.180</td>
<td>192.805</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Huynh-Feldt</td>
<td>20279.273</td>
<td>110.254</td>
<td>183.932</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lower-bound</td>
<td>20279.273</td>
<td>58.000</td>
<td>349.643</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 10.6 Pairwise comparison of the accuracy scores on gender referents in the metalinguistic knowledge test

Tests of Within-Subjects Contrasts

<table>
<thead>
<tr>
<th>Source</th>
<th>Level 1 vs. Level 2</th>
<th>Level 2 vs. Level 3</th>
<th>Level 1 vs. Level 2</th>
<th>Level 2 vs. Level 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>time</td>
<td>2086.987</td>
<td>158.072</td>
<td>2086.987</td>
<td>158.072</td>
</tr>
<tr>
<td></td>
<td>4.583</td>
<td>.598</td>
<td>.037</td>
<td>.443</td>
</tr>
<tr>
<td>time * group</td>
<td>246.987</td>
<td>98.072</td>
<td>246.987</td>
<td>98.072</td>
</tr>
<tr>
<td></td>
<td>.542</td>
<td>.371</td>
<td>.464</td>
<td>.545</td>
</tr>
<tr>
<td>Error(time)</td>
<td>26411.346</td>
<td>15335.261</td>
<td>455.368</td>
<td>264.401</td>
</tr>
</tbody>
</table>
Table 10.7 ANOVA result for Third-person -s in the elicited oral imitation test

<table>
<thead>
<tr>
<th></th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>TPMpre</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Between Groups</td>
<td>80.208</td>
<td>1</td>
<td>80.208</td>
<td>.235</td>
<td>.630</td>
</tr>
<tr>
<td>Within Groups</td>
<td>19813.126</td>
<td>58</td>
<td>341.606</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>19893.333</td>
<td>59</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TPMpost</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Between Groups</td>
<td>11624.056</td>
<td>1</td>
<td>11624.056</td>
<td>42.808</td>
<td>.000</td>
</tr>
<tr>
<td>Within Groups</td>
<td>15749.277</td>
<td>58</td>
<td>271.539</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>27373.333</td>
<td>59</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TPMdpt</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Between Groups</td>
<td>11614.774</td>
<td>1</td>
<td>11614.774</td>
<td>43.173</td>
<td>.000</td>
</tr>
<tr>
<td>Within Groups</td>
<td>15603.560</td>
<td>58</td>
<td>269.027</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>27218.333</td>
<td>59</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 10.8  ANOVA result for Third-person -s in the metalinguistic knowledge test

<table>
<thead>
<tr>
<th></th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>TPMPre</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Between Groups</td>
<td>66.224</td>
<td>1</td>
<td>66.224</td>
<td>.486</td>
<td>.488</td>
</tr>
<tr>
<td>Within Groups</td>
<td>7898.776</td>
<td>58</td>
<td>136.186</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>7965.000</td>
<td>59</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TPMpost</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Between Groups</td>
<td>656.331</td>
<td>1</td>
<td>656.331</td>
<td>5.286</td>
<td>.025</td>
</tr>
<tr>
<td>Within Groups</td>
<td>7202.002</td>
<td>58</td>
<td>124.172</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>7858.333</td>
<td>59</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TPMdpt</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Between Groups</td>
<td>660.751</td>
<td>1</td>
<td>660.751</td>
<td>7.944</td>
<td>.007</td>
</tr>
<tr>
<td>Within Groups</td>
<td>4824.249</td>
<td>58</td>
<td>83.177</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>5485.000</td>
<td>59</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 10.9 Repeated measures ANOVA result for third-person -s in the elicited oral imitation test

### Tests of Within-Subjects Effects

<table>
<thead>
<tr>
<th>Source</th>
<th>Type III Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>time</td>
<td>Sphericity Assumed</td>
<td>26699.237</td>
<td>2</td>
<td>13349.619</td>
<td>49.182</td>
</tr>
<tr>
<td></td>
<td>Greenhouse-Geisser</td>
<td>26699.237</td>
<td>1.923</td>
<td>13886.810</td>
<td>49.182</td>
</tr>
<tr>
<td></td>
<td>Huynh-Feldt</td>
<td>26699.237</td>
<td>2.000</td>
<td>13349.619</td>
<td>49.182</td>
</tr>
<tr>
<td></td>
<td>Lower-bound</td>
<td>26699.237</td>
<td>1.000</td>
<td>26699.237</td>
<td>49.182</td>
</tr>
<tr>
<td>time * group</td>
<td>Sphericity Assumed</td>
<td>6512.571</td>
<td>2</td>
<td>3256.285</td>
<td>11.997</td>
</tr>
<tr>
<td></td>
<td>Greenhouse-Geisser</td>
<td>6512.571</td>
<td>1.923</td>
<td>3387.319</td>
<td>11.997</td>
</tr>
<tr>
<td></td>
<td>Huynh-Feldt</td>
<td>6512.571</td>
<td>2.000</td>
<td>3256.285</td>
<td>11.997</td>
</tr>
<tr>
<td></td>
<td>Lower-bound</td>
<td>6512.571</td>
<td>1.000</td>
<td>6512.571</td>
<td>11.997</td>
</tr>
<tr>
<td>Error(time)</td>
<td>Sphericity Assumed</td>
<td>31486.318</td>
<td>116</td>
<td>271.434</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Greenhouse-Geisser</td>
<td>31486.318</td>
<td>111.513</td>
<td>282.356</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Huynh-Feldt</td>
<td>31486.318</td>
<td>116.000</td>
<td>271.434</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Lower-bound</td>
<td>31486.318</td>
<td>58.000</td>
<td>542.868</td>
<td></td>
</tr>
</tbody>
</table>

### Tests of Between-Subjects Effects

<table>
<thead>
<tr>
<th>Source</th>
<th>Type III Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>160264.378</td>
<td>1</td>
<td>160264.378</td>
<td>1416.997</td>
<td>.000</td>
</tr>
<tr>
<td>group</td>
<td>5602.156</td>
<td>1</td>
<td>5602.156</td>
<td>49.532</td>
<td>.000</td>
</tr>
<tr>
<td>Error</td>
<td>6559.881</td>
<td>58</td>
<td>113.101</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 10.10 Pairwise comparison of the accuracy scores on third person -s in the elicited oral imitation test

<table>
<thead>
<tr>
<th>Source</th>
<th>time * group</th>
<th>Type III Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>time</td>
<td>Level 1 vs. Level 2</td>
<td>31959.778</td>
<td>1</td>
<td>31959.778</td>
<td>49.829</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td>Level 2 vs. Level 3</td>
<td>1400.002</td>
<td>1</td>
<td>1400.002</td>
<td>3.088</td>
<td>.084</td>
</tr>
<tr>
<td>time * group</td>
<td>Level 1 vs. Level 2</td>
<td>9773.111</td>
<td>1</td>
<td>9773.111</td>
<td>15.238</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td>Level 2 vs. Level 3</td>
<td>.002</td>
<td>1</td>
<td>.002</td>
<td>.000</td>
<td>.998</td>
</tr>
<tr>
<td>Error(time)</td>
<td>Level 1 vs. Level 2</td>
<td>37200.222</td>
<td>58</td>
<td>641.383</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Level 2 vs. Level 3</td>
<td>26298.331</td>
<td>58</td>
<td>453.420</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 10.11 Repeated measures ANOVA result for third-person -s in the metalinguistic knowledge test

**Tests of Within-Subjects Effects**

<table>
<thead>
<tr>
<th>Source</th>
<th>Type III Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>time</td>
<td>Sphericity Assumed</td>
<td>2502.514</td>
<td>2</td>
<td>1251.257</td>
<td>13.951</td>
</tr>
<tr>
<td></td>
<td>Greenhouse-Geisser</td>
<td>2502.514</td>
<td>1.700</td>
<td>1472.166</td>
<td>13.951</td>
</tr>
<tr>
<td></td>
<td>Huynh-Feldt</td>
<td>2502.514</td>
<td>1.776</td>
<td>1409.015</td>
<td>13.951</td>
</tr>
<tr>
<td></td>
<td>Lower-bound</td>
<td>2502.514</td>
<td>1.000</td>
<td>2502.514</td>
<td>13.951</td>
</tr>
<tr>
<td>time * group</td>
<td>Sphericity Assumed</td>
<td>204.736</td>
<td>2</td>
<td>102.368</td>
<td>1.141</td>
</tr>
<tr>
<td></td>
<td>Greenhouse-Geisser</td>
<td>204.736</td>
<td>1.700</td>
<td>120.441</td>
<td>1.141</td>
</tr>
<tr>
<td></td>
<td>Huynh-Feldt</td>
<td>204.736</td>
<td>1.776</td>
<td>115.275</td>
<td>1.141</td>
</tr>
<tr>
<td></td>
<td>Lower-bound</td>
<td>204.736</td>
<td>1.000</td>
<td>204.736</td>
<td>1.141</td>
</tr>
<tr>
<td>Error(time)</td>
<td>Sphericity Assumed</td>
<td>10404.153</td>
<td>116</td>
<td>89.691</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Greenhouse-Geisser</td>
<td>10404.153</td>
<td>98.593</td>
<td>105.526</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Huynh-Feldt</td>
<td>10404.153</td>
<td>103.012</td>
<td>100.999</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Lower-bound</td>
<td>10404.153</td>
<td>58.000</td>
<td>179.382</td>
<td></td>
</tr>
</tbody>
</table>

**Tests of Between-Subjects Effects**

Measure: MEASURE_1
Transformed Variable: Average

<table>
<thead>
<tr>
<th>Source</th>
<th>Type III Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>505155.819</td>
<td>1</td>
<td>505155.819</td>
<td>9232.041</td>
<td>.000</td>
</tr>
<tr>
<td>group</td>
<td>392.856</td>
<td>1</td>
<td>392.856</td>
<td>7.180</td>
<td>.010</td>
</tr>
<tr>
<td>Error</td>
<td>3173.625</td>
<td>58</td>
<td>54.718</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 10.12 Pairwise comparison of the accuracy scores on third person -s in the metalinguistic test

**Tests of Within-Subjects Contrasts**

<table>
<thead>
<tr>
<th>Source</th>
<th>time</th>
<th>Type III Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>time</td>
<td>Level 1 vs. Level 2</td>
<td>3592.258</td>
<td>1</td>
<td>3592.258</td>
<td>14.109</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td>Level 2 vs. Level 3</td>
<td>6.674</td>
<td>1</td>
<td>6.674</td>
<td>.048</td>
<td>.827</td>
</tr>
<tr>
<td>time *</td>
<td>Level 1 vs. Level 2</td>
<td>305.591</td>
<td>1</td>
<td>305.591</td>
<td>1.200</td>
<td>.278</td>
</tr>
<tr>
<td>group</td>
<td>Level 2 vs. Level 3</td>
<td>.007</td>
<td>1</td>
<td>.007</td>
<td>.000</td>
<td>.994</td>
</tr>
<tr>
<td>Error</td>
<td>Level 1 vs. Level 2</td>
<td>14767.742</td>
<td>58</td>
<td>254.616</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(time)</td>
<td>Level 2 vs. Level 3</td>
<td>7993.326</td>
<td>58</td>
<td>137.816</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
BIBLIOGRAPHY


CURRICULUM VITAE

QIN WANG

School of English Studies
Dalian University of Foreign Languages
Lvshun South Road West End 6, Dalian, P.R.C, 116044
melodyqw@126.com

EDUCATION:

Boston University, Boston, USA 2016
Doctor of Education in Developmental Studies
Area of Concentration: Language Education

Boston University, Boston, USA 2002
Master of Education in Curriculum and Teaching
Area of Concentration: TESOL

Liao Ning Normal University, Dalian, China 1995
Bachelor of Arts in English Language and Literature

PROFESSIONAL EXPERIENCE:

Dalian University of Foreign Languages, Dalian, China 2003–present
English lecturer and associate professor

BBN Technology, Boston, USA 2008–2009
Language annotator

Boston University, Boston, USA 2007–2009
Chinese instructor

Boston Chinatown Neighborhood Center, Boston, USA 2003
English instructor

Dalian University, Dalian, China 1995–2001
English Lecturer
SELECTED PUBLICATIONS:


GRANTS:

Foreign Language Teaching and Reform Grant, Liaoning Province Higher Education Association, 2014

University Teaching Reform Grant, Dalian University of Foreign Languages, 2013

University Research Grant, Dalian University of Foreign Languages, 2012

Social Science Research Grant, Liaoning Province Social Science Planning Office, 2012

University Teaching Reform Grant, Dalian University of Foreign Languages, 2007

Province Higher Education Research Grant, Liaoning Province Department of Education, 2005

PROFESSIONAL MEMBERSHIPS:

TESOL International Association

Pi Lambda Theta

AWARDS:

Graduate Research and Scholarship Award, Boston University School of Education, 2013


Paper of Excellence, the 14th EFL Conference in Northeast China, 2006

Third Prize Winner of the University Research Achievements, Dalian University of Foreign Languages, 2006