THE EFFECT OF IMPLICIT AND EXPLICIT FEEDBACK: A STUDY ON THE ACQUISITION OF MANDARIN CLASSIFIERS BY CHINESE HERITAGE AND NON-HERITAGE LANGUAGE LEARNERS

By

YE HAN

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

2010
To Jiuru Han and Yulan Zhou
ACKNOWLEDGMENTS

First of all, I own my deepest gratitude to my dissertation committee advisor Theresa Antes for her encouragement, supervision and support throughout the time it took me to complete the project. I thank the members of my dissertation committee, Diana Boxer, Brent Henderson, and Ester De Jong, who have generously contributed their time and expertise to improve my work. I would also like to thank Takako Egi. It would not be possible to complete the project without her inspiration and guidance at the preliminary stage of my dissertation project.

I am grateful for the faculty and teaching assistants in the Chinese program, Cynthia Chennault, Cynthia Shen, Elinore Fresh, Han Xu, I-Chun Peir, Richard Wang, Shizhong Deng, Binmei Liu, Jimmy Huang, Rui Cao, Ya-Gang Fan, and Yeng Jung Lee. I also thank all the students who participated in my research. I would also like to acknowledge Janet Rose and Judy Shoaf for their help in arranging the data collection location and equipment.

I would also like to thank all my fellow graduate school friends, for their encouragement and support. My thanks must go also to my research assistant, Jingya Zhong, who spent many hours assisting in data coding.

I particularly want to thank my parents, who have supported me in every respect during my five years in graduate school. I thank my father, who inspired me to develop a life-long passion toward books and knowledge by consistent perusing higher levels of achievement in his own field throughout his entire career. I thank my mother, who inspired me to become a brave and strong person, who is not afraid of challenges and obstacles in life.
# TABLE OF CONTENTS

**ACKNOWLEDGMENTS** .......................................................................................................................... 4

**LIST OF TABLES** ................................................................................................................................. 8

**LIST OF FIGURES** ............................................................................................................................... 11

**ABSTRACT** .......................................................................................................................................... 12

**CHAPTER**

1 **INTRODUCTION** ............................................................................................................................... 14

2 **HERITAGE LANGUAGE LEARNERS** ............................................................................................... 18

   2.1 Who Are Heritage Language Learners? ......................................................................................... 18
      2.1.1 Defining Heritage Language learners .................................................................................. 18
      2.1.2 Heritage Language Learners’ Language Proficiency .......................................................... 21
      2.1.3 Predictive Variables of Heritage Language Learners’ Proficiency In Their Heritage Language .......................................................... 24
      2.1.4 Research on Similarities and Differences between Heritage Language and Non-Heritage Language Learners ........................................... 27

   2.2 Chinese Heritage Language Learners ............................................................................................ 34
      2.2.1 Chinese Dialects .................................................................................................................... 34
      2.2.2 Chinese People in the U.S. .................................................................................................... 36
      2.2.3 Defining Chinese Heritage Language Learners .................................................................. 38
      2.2.4 The Differences between Chinese Heritage Language and Non-Chinese Heritage Language Learners .......................................................... 42

3 **IMPLICIT AND EXPLICIT FEEDBACK** ....................................................................................... 50

   3.1 Theoretical Constructs ................................................................................................................... 50
      3.1.1 Implicit vs. Explicit Learning ................................................................................................. 50
      3.1.2 Implicit vs. Explicit Knowledge ........................................................................................... 53
      3.1.3 Implicit vs. Explicit Feedback .............................................................................................. 56

   3.2 Measuring Effectiveness of Corrective Feedback ......................................................................... 62
      3.2.1 Uptake and Learner Repair .................................................................................................. 62
      3.2.2 Private Speech ...................................................................................................................... 64
      3.2.3 Learner’ Noticing of Feedback ............................................................................................ 66
      3.2.4 Test Performance ................................................................................................................ 72

   3.3 Previous Empirical Research Comparing the Effects of Implicit and Explicit Feedback ............. 75

   3.4 Chinese Heritage Language Learners, Implicit and Explicit Feedback ....................................... 93
4 RESEARCH QUESTIONS AND METHODOLOGY ................................................. 101

4.1 Research Questions .................................................................................. 101
4.2 Methodology .............................................................................................. 103
  4.2.1 Operationalizations ............................................................................. 103
  4.2.2 Design .................................................................................................. 106
  4.2.3 Participants ......................................................................................... 109
  4.2.4 Target Linguistic Items ....................................................................... 116
  4.2.5 Materials .............................................................................................. 120
    4.2.5.1 Background information questionnaire ........................................ 120
    4.2.5.2 Treatment sessions ........................................................................ 122
    4.2.5.3 Testing material .............................................................................. 124
    4.2.5.4 The oral imitation test ..................................................................... 127
    4.2.5.5 The untimed written cloze test ...................................................... 129
    4.2.5.6 Advantages and disadvantages of the testing design in the current study .... 130
    4.2.5.7 Measurement of learners’ perception of the feedback ................. 131
    4.2.5.8 Interview ......................................................................................... 135
  4.2.6 Procedure .............................................................................................. 136
  4.2.7 Transcriptions ....................................................................................... 139
  4.2.8 Scoring .................................................................................................. 139
  4.2.9 Coding ................................................................................................... 139
    4.2.9.1 Classifier-related comments .......................................................... 140
    4.2.9.2 Non classifier-related comments ................................................... 143
  4.2.10 Inter-Rater Reliability ......................................................................... 146

5 RESULTS ......................................................................................................... 148

5.1 Research Question 1 ................................................................................ 148
  5.1.1 Descriptive Statistics .......................................................................... 148
    5.1.1.1 Classifier-related comments ......................................................... 150
    5.1.1.2 Non classifier-related comments ................................................... 152
  5.1.2 Inferential Statistics ............................................................................. 153
5.2 Research Question 2 ................................................................................ 154
  5.2.1 Descriptive Statistics .......................................................................... 154
    5.2.1.1 Oral test ......................................................................................... 154
    5.2.1.2 Written test .................................................................................... 156
    5.2.1.3 Learners’ previous knowledge reflected in their verbal comments ............................................................................ 159
  5.2.2 Inferential Statistics ............................................................................. 161
5.3 Research Question 3 ................................................................................ 164
  5.3.1 Descriptive Statistics .......................................................................... 164
    5.3.1.1 Oral imitation test ........................................................................... 164
    5.3.1.2 Written cloze test ........................................................................... 165
  5.3.2 Inferential Statistics ............................................................................. 167
6 DISCUSSION AND CONCLUSION ................................................................................. 169

6.1 Research Question 1 ................................................................................................. 169
6.2 Research Question 2 ................................................................................................. 185
6.3 Research Question 3 ................................................................................................. 193
6.4 Implications ............................................................................................................... 207
  6.4.1 Theoretical Implications ..................................................................................... 207
  6.4.2 Pedagogical Implications .................................................................................... 210
6.5 Limitations and Future Research .............................................................................. 212
6.6 Conclusion ................................................................................................................. 215

APPENDIX

A TRANSCRIPTION CONVENTIONS ............................................................................. 217
B CONSENT FORM ........................................................................................................... 218
C QUESTIONNAIRE ........................................................................................................ 220
D WRITTEN PRE-TEST .................................................................................................... 222
E WRITTEN POST-TEST ................................................................................................. 224
F TREATMENT 1 .............................................................................................................. 226
G TREATMENT 2 .............................................................................................................. 227
LIST OF REFERENCES ................................................................................................... 228
BIOGRAPHICAL SKETCH ............................................................................................... 239
<table>
<thead>
<tr>
<th>Table</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3-1</td>
<td>Key characteristics of implicit and explicit knowledge</td>
<td>54</td>
</tr>
<tr>
<td>3-2</td>
<td>Laboratory studies of effects of implicit and explicit feedback</td>
<td>87</td>
</tr>
<tr>
<td>3-2</td>
<td>Continued</td>
<td>88</td>
</tr>
<tr>
<td>3-3</td>
<td>Classroom studies of effects of implicit and explicit feedback</td>
<td>89</td>
</tr>
<tr>
<td>3-3</td>
<td>Continued</td>
<td>90</td>
</tr>
<tr>
<td>3-3</td>
<td>Continued</td>
<td>91</td>
</tr>
<tr>
<td>3-4</td>
<td>Computer assisted language learning studies of effects of implicit and explicit feedback</td>
<td>92</td>
</tr>
<tr>
<td>4-1</td>
<td>The targeted classifiers</td>
<td>121</td>
</tr>
<tr>
<td>4-2</td>
<td>The experimental and control groups</td>
<td>137</td>
</tr>
<tr>
<td>5-1</td>
<td>Feedback episodes provided for Chinese heritage language learner group</td>
<td>148</td>
</tr>
<tr>
<td>5-2</td>
<td>Feedback episodes provided for non-Chinese heritage language learner group</td>
<td>149</td>
</tr>
<tr>
<td>5-3</td>
<td>Commented feedback episodes by Chinese heritage language learner group</td>
<td>150</td>
</tr>
<tr>
<td>5-4</td>
<td>Commented feedback episodes by non-Chinese heritage language learner group</td>
<td>150</td>
</tr>
<tr>
<td>5-5</td>
<td>Classifier-related comments provided by Chinese heritage language learner group</td>
<td>151</td>
</tr>
<tr>
<td>5-6</td>
<td>Classifier-related comments provided by non-Chinese heritage language learner group</td>
<td>151</td>
</tr>
<tr>
<td>5-7</td>
<td>Chinese heritage language learners' perceptions of feedback not identified as classifier-related</td>
<td>153</td>
</tr>
<tr>
<td>5-8</td>
<td>Non-Chinese heritage language learners’ perceptions of feedback not identified as classifier-related</td>
<td>153</td>
</tr>
<tr>
<td>5-9</td>
<td>Summary of results of LSD post hoc test</td>
<td>154</td>
</tr>
<tr>
<td>5-10</td>
<td>Summary of oral pre-test scores for Chinese heritage language learner group</td>
<td>155</td>
</tr>
</tbody>
</table>
5-11 Summary of oral pre-test scores for non-Chinese heritage language learner group ................................................................. 155

5-12 Summary of oral post-test scores of Chinese heritage language learner group .................................................................................. 156

5-13 Summary of oral post-test scores of non-Chinese heritage language learner group ................................................................................. 156

5-14 Summary of written pre-test scores for Chinese heritage language learner group .................................................................................. 157

5-15 Summary of written pre-test scores for non-Chinese heritage language learner group ................................................................................. 157

5-16 Summary of written post-test scores for Chinese heritage language learner group .................................................................................. 158

5-17 Summary of written post-test scores for non-Chinese heritage language learner group ................................................................................. 158

5-18 Comparison of pre-test and post-test scores across the groups .............. 159

5-19 Chinese heritage language learners’ previous knowledge of target classifiers 160

5-20 Non-Chinese heritage language learners’ previous knowledge of target classifiers .................................................................................. 160

5-21 One-way ANOVA for the oral pre-test ................................................................. 161

5-22 One-way ANOVA for the written pre-test ................................................................. 161

5-23 One-way ANOVA for the oral post-test ................................................................. 162

5-24 One-way ANOVA for the written post-test ................................................................. 162

5-25 LSD post hoc test for Chinese heritage language learners on the oral test .... 162

5-26 LSD post hoc test for non-Chinese heritage language learners on the oral test .................................................................................. 162

5-27 LSD post hoc test for Chinese heritage language learners on the written test . 163

5-28 LSD post hoc test for non-Chinese heritage language learners on the written test .................................................................................. 163

5-29 Summary of increase on the oral post-test for recast group ......................... 164

5-30 Summary of increase on the oral post-test for meta-linguistic group .......... 164
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-1</td>
<td>A continuum of heritage language learners’ proficiency</td>
<td>22</td>
</tr>
<tr>
<td>3-1</td>
<td>The explicitness of the feedback</td>
<td>61</td>
</tr>
<tr>
<td>4-1</td>
<td>The research design</td>
<td>108</td>
</tr>
<tr>
<td>4-2</td>
<td>Experimental procedure and approximate time</td>
<td>138</td>
</tr>
<tr>
<td>4-3</td>
<td>Coding scheme of the recall data</td>
<td>140</td>
</tr>
<tr>
<td>5-1</td>
<td>Feedback episodes provided for Chinese heritage language group</td>
<td>149</td>
</tr>
<tr>
<td>5-2</td>
<td>Feedback episodes provided for non-Chinese heritage language group</td>
<td>149</td>
</tr>
<tr>
<td>5-3</td>
<td>Percentage of classifier-related comments across experimental groups</td>
<td>151</td>
</tr>
<tr>
<td>5-4</td>
<td>Mean scores of oral pre-test</td>
<td>155</td>
</tr>
<tr>
<td>5-5</td>
<td>Mean scores of oral post-test</td>
<td>156</td>
</tr>
<tr>
<td>5-6</td>
<td>Mean scores of written pre-test</td>
<td>157</td>
</tr>
<tr>
<td>5-7</td>
<td>Mean scores of written post-test</td>
<td>158</td>
</tr>
<tr>
<td>5-8</td>
<td>Chinese heritage language learners’ previous knowledge of target classifiers</td>
<td>160</td>
</tr>
<tr>
<td>5-9</td>
<td>Non-Chinese heritage language learners’ previous knowledge of target classifiers</td>
<td>160</td>
</tr>
<tr>
<td>5-10</td>
<td>Increase on the oral post-test across the experimental groups</td>
<td>165</td>
</tr>
<tr>
<td>5-11</td>
<td>Increases on the written post-test across the experimental groups</td>
<td>166</td>
</tr>
</tbody>
</table>
Abstract of Dissertation Presented to the Graduate School of the University of Florida in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

THE EFFECT OF IMPLICIT AND EXPLICIT FEEDBACK: A STUDY ON THE ACQUISITION OF MANDARIN CLASSIFIERS BY CHINESE HERITAGE AND NON-HERITAGE LANGUAGE LEARNERS

By

Ye Han

August 2010

Chair: Theresa Antes
Major: Linguistics

Previous studies revealed mixed results in terms of the relative effects of implicit and explicit feedback: some found that explicit feedback worked more efficiently than implicit feedback; others found no difference between the two feedback types. These contrasting results called for further investigations into this issue, particularly examining those underexplored learner groups who have developed their target language (TL) in a different path from traditional foreign/second language learners investigated in previous studies, such as heritage language (HL) learners. Therefore the present study aims at contributing to the literature by investigating the relative effects of implicit feedback in the form of recasts and explicit feedback in the form of meta-linguistic feedback on the acquisition of Mandarin classifiers by Chinese heritage language (CHL) and non-heritage language (non-CHL) learners. This study employed a pre-test-treatment-post-test research design, in which 64 participants (CHL=35, non-CHL=29) were randomly divided into four experimental groups and two control groups. Feedback was provided on learners’ errors by a native Chinese interlocutor during the two treatment sessions. Learners’ perceptions of feedback were measured by their verbal comments provided
during the stimulated recall, and their acquisition of classifiers were assessed by their test scores. The data were both quantitatively and qualitatively analyzed. The results showed that both feedback types were effective in facilitating learners’ acquisition. In addition, learners’ language background was also found to affect their perceptions of feedback, as well as their acquisition. The findings of this study expand our knowledge about implicit and explicit feedback. In addition, they also provide invaluable information particularly for educators and administrators who are involved in HL instruction.
CHAPTER 1
INTRODUCTION

The present study examines the effects of implicit and explicit feedback on the acquisition of Mandarin classifiers by learners from heritage language (HL) and non-heritage language (non-HL) backgrounds. The study is rooted within an interactionist framework, drawing from previous research in second language acquisition (SLA) on the topic. Implicit feedback is operationalized in the form of recasts, which are defined as a native speakers’ (NS) reformulation of learners’ problematic utterances; explicit feedback takes the form of meta-linguistic feedback, which refers to a provision of meta-linguistic information to indicate an error made by the learner through highlighting the nature and the characteristic of the target language (TL) form.

Previous research on the effects of implicit and explicit feedback has revealed controversial results: some researchers found that implicit and explicit feedback worked equally effectively (e.g., Carroll, 2001; DeKeyser, 1993; Kang, 2009; Kim & Mathes, 2001; Leeman, 2003; Loewen & Nabei, 2007), while others reported the superiority of explicit feedback over implicit feedback (e.g., Ammar & Spada, 2006; Lyster, 2004; Murano, 2000; R. Ellis et al., 2006). In previous studies, learners usually studied the TL as a foreign language (FL) or a second language (SL) through formal classroom instruction; therefore they had developed their explicit knowledge and literacy skills, but were weak in their implicit knowledge. However, little empirical research has been done to investigate learners who have developed implicit knowledge but lack explicit knowledge and literacy skills due to their exposure during childhood, such as HL learners. In addition, in the United States, HL instruction and research has been attracting more and more interest from researchers and language educators, along with
the rapid growth of the HL population in postsecondary FL/ SL classrooms since the 1970s (Valdés, 2005). In less than two decades, the field of CHL instruction has made astounding progress. More and more language teachers in postsecondary Chinese programs have to face CHL learners and non-CHL learners in the same classroom. Due to their different language backgrounds in Chinese, these two groups of learners have different language behaviors and language needs. However, little empirical research has been done in the past to compare the linguistic behaviors of these two groups of learners.

Motivated by gaps in the literature, the present study empirically investigates the effects of recasts and meta-linguistic feedback on learners with CHL background, in comparison with learners without CHL background. The following three issues are particularly explored: learners’ perceptions of recasts and meta-linguistic feedback; the developmental benefits of recasts and meta-linguistic feedback; the relationship between feedback type and the increase in learners’ performance on the oral and written tests.

Chapter 2 highlights some of the fundamental issues associated with HL learners, with a particular emphasis on CHL learners. In particular, it asks and answers the following questions: who are HL learners? How do they differ from non-HL learners? In the first part, the chapter will examine HL learners in general. It will provide definitions of HL learners based on previous studies. Then it will discuss issues related to HL learners’ language proficiency, including various variables that could be employed to predict their proficiency levels. The chapter will then present a review of previous research on the comparison between HL learners and non-HL learners. The second
part of the chapter will focus on CHL learners in particular. It starts with an overview of Chinese language and Chinese immigrants in the U.S., followed by a review of definitions of CHL learners provided in the literature. Finally, the chapter will discuss the differences between CHL learners and non-CHL learners in linguistic, cultural and social aspects based on the review of previous research.

Chapter 3 will discuss effects of implicit and explicit feedback within an interactionist framework. The chapter will first introduce some theoretical constructs which relate to the research goal of the present study, including implicit and explicit learning; implicit and explicit knowledge; and implicit and explicit feedback. Then the chapter will introduce four measures that have been frequently employed in previous studies to assess the effectiveness of corrective feedback, such as uptake and learner repair, private speech, learners’ noticing of feedback, and test performance. The chapter will then review previous empirical research on the comparison of the effects of implicit and explicit feedback. Finally, the chapter will provide a working definition of implicit feedback and explicit feedback for the present study, based on a discussion of CHL learners’ profiles and internal knowledge systems.

Based on the theoretical basis built from the literature review in Chapters 2 and 3, Chapter 4 will present research questions that were motivated by gaps in the previous research, as well as the methodology employed to answer the research questions. It will describe the materials used to collect the data in the study, as well as the analytical tests employed to evaluate the data. The results of the experiment will be presented in Chapter 5. The chapter will include three sections, each of which will report one
research question. For each research question, the results of descriptive statistics will be reported first, followed by the results of inferential statistics.

In Chapter 6, the findings will be discussed with regard to the research questions. I will first suggest some explanations for the findings about the learners’ perceptions about meta-linguistic feedback and recasts (Research Question 1), and then discuss the relative effects of the two types of feedback (Research Question 2). Finally, I will discuss the findings of the relationship between feedback type and learners’ increase in their performance on the post-tests. Learners’ language background and feedback type will be considered as the two most important independent variables throughout the whole discussion. Following the discussion of the findings, the theoretical and pedagogical implications and some limitations of the current study will be discussed. Possibilities of future research will be raised. The chapter concludes with a summary of how the findings of this study can extend our knowledge of effectiveness of feedback on learners with different language backgrounds.
CHAPTER 2
HL LEARNERS

2.1 Who Are HL Learners?

2.1.1 Defining HL learners

In the U.S., before the term *HL learner* was introduced in the FL teaching profession in the mid 1990s, HL learners were labeled as native speakers, quasi-native speakers, or bilingual students. However, none of these terms is able to effectively distinguish this group of learners from traditionally acknowledged learner groups in the literature, such as an L1 learner group, or an L2 learner group. In fact, HL learners share both similarities and differences with L1 and L2 learners: like L1 learners, they experience early exposure to the HL and culture “in a naturalistic setting during a critical period for the development of language” (Montrul 2008, p.500). However, these exposures are usually either restricted or insufficient. Therefore their knowledge and ability of the HL are usually incomplete, which is similar to L2 learners (Carreira, 2004; Egi, 2007; Gregg, 1984; Han, 2000; Koda et al., 2008a; Loewen, 2004; McLaughlin, 1978; Montrul, 2008; Russo et al., 1989; Sanz et al., 2009; Sauro, 2009; Slimani-Rolls, 2005; Valdés, 1995; 2001; Wu, 2008). On the other hand, their historical, family and ethnic connections to their HL are not shared by L2 learners, and differ from L1 learners. The term *HL learner* highlighted this unique connection.

Since HL learners have unique language experiences and linguistic abilities which set them apart from the traditional L1, L2 and FL learners, in the language education field, it is important to clarify who are the HL learners, particularly for those who work on curriculum design, materials selection, student placement and assessment, and teacher training (Carreira, 2004). However, this group consists of people with various historical,
social, linguistic, and demographic backgrounds: they can be foreign-born immigrants who moved to the U.S. at a young age, native-born children with foreign-born immigrant parents, or native-born children of native-born parents with immigrant background (Valdés, 2005). Because of the heterogeneous nature of the population in this group, defining the HL learner is not an easy task. In the literature, three criteria have been employed to define HL learners: their membership in the HL community, their personal connection through family background, and their greater or lesser linguistic proficiency in the HL (Carreira, 2004). Based on the above criteria, numerous definitions have been proposed. These definitions can be roughly divided into two groups: a *broadly defined group*, and a *narrowly defined group* (Kondo-Brown, 2005).

Definitions belonging to the *broadly defined group* tend to focus on learners’ ethnic background connection to their “own ancestral language (i.e., indigenous, colonial, and immigrant languages)” (Kondo-Brown, 2005, p.564), or their “heritage motivation” (Van Deusen-Scholl, 2003, p.222). The most influential and representative definition in this group proposed by Fishman, defined HL learners based on two criteria: learners who speak LOTEs (languages other than English), and who have a particular family relevance to the language (2001).

Since definitions in the *broadly defined group* only emphasize the ancestral, ethnic background connection to the language, without considering proficiency level, they often include within the group many members who actually cannot speak a word in the HL (Van Deusen-Scholl, 2003). Consequently, in university-level FL classes, students who seek to “reconnect with their family’s heritage through language” are also considered as HL learners, “even though the linguistic evidence of that connection may have been lost
for generations” (Van Deusen-Scholl, 2003, p. 222). Thus definitions in the *broadly defined group* provide language educators with little help in distinguishing HL learners from traditional FL/L2 learners in the classroom.

On the other hand, definitions in the *narrowly defined group* adopt language proficiency and language affiliation as the criteria to restrict the members within the HL learner group. Language affiliation refers to “degree of personal or emotional commitment or connection to the home, community, or other languages and cultures” (Li & Duff, 2008, p.16).

The most widely accepted definition has been proposed by Valdés. In her definition, a HL learner refers to someone who was “raised in a home where a non-English language is spoken” and who “speaks or at least understands the language and who is to some degree bilingual in that language and in English” (2001, p.38). In this definition, Valdés sets three main criteria for identifying HL learners: the home language, minimal proficiency in the HL, and the societal language. This definition restricts HL learners to those who have achieved some degree of bilingual proficiency (Kondo-Brown, 2005), excluding individuals with strong family or personal connections to the HL but who cannot speak the language.

Nevertheless, the proficiency of learners is defined in a relatively broad and vague manner in Valdés’s definition. A similar but more detailed description of learners’ proficiency level was given in the National Standards for Foreign Language Education Project (1999): “these students may come to class able to converse in the language in home and community situations but may lack the abilities to interact comfortably in more formal settings. Further, they may be quite comfortable with oral language but possess
limited skills in reading and writing” (1999, p. 29). This description highlighted at least two characteristics of HL learners’ proficiency: they tend to have stronger oral proficiency than reading and writing abilities; in terms of register, they are more likely to control informal than formal registers. However, fully assessing the language proficiency of a HL learner is a far more complex issue, which I will address in the next section.

2.1.2 HL Learners’ Language Proficiency

Although HL learners grow up in a bilingual (or even multilingual) environment, they hardly have “access to two languages in exactly the same contexts in every domain of interaction,” and rarely use both languages to “carry out the exact same functions with every person with whom they interact” (Valdés, 2005, p.414). In fact, HL learners mostly use HL only with their family members (e.g., parents, grandparents, etc), and people from the communities where the HL is spoken (e.g., church, language school). They usually receive their formal education at mainstream schools where the entire curriculum is conducted in English. Unless being taught at home or sent to after-school community HL schools, they usually develop their literacy only in English (Valdés, 1995; 2001). Therefore HL learners are not full bilinguals or balanced bilinguals who have absolutely equivalent proficiencies in two languages. Instead, they are bilingual individuals “who manifest very different strengths in their two languages” (Valdés, 2005, p.414). Their proficiency can be considered within a “continuum of linguistic abilities and communicative strategies that an individual may access in one or the other of his or her two languages at a specific moment, for a particular purpose, in a particular setting, with particular interlocutors” (Valdés, 1995, p.316), as illustrated in Figure 2-1.
In Figure 2-1, the size of the fonts indicates language strengths in languages A and B respectively. The two letters --“A” and “B”--at the left and right extremes, indicate monolinguals who speak only language A or B. The two same-sized letters “AB” in the middle indicate absolute equivalent language proficiency in languages A and B, which is only theoretically possible. As illustrated in Figure 2-1, due to limited linguistic input, HL learners usually never acquire all the domains in their HL completely, such as grammatical, textual, illocutionary, and sociolinguistic competence (Bachman, 1990). Therefore HL learners usually have unbalanced abilities in their HL and societal language, and this unbalance usually will be further expanded when the “type and quantity of linguistic input and linguistic interaction necessary to maintain the full lexical, phonological, morphological, and syntactic distinctions that are made by fluent competent speakers of this language” (Anderson, 1982, p.91) become unavailable or less available. Previous studies found that immigrant children rapidly lose their HL skills after starting their schooling in mainstream schools, and their dominant language will quickly shift from HL to English (Fishman, 1991; Wong Fillmore, 1991). Some of them may lose their HL completely; others may try to reconnect with their HL through formal education, such as taking language courses at universities or colleges. However, even

<table>
<thead>
<tr>
<th>Monolingual</th>
<th>Monolingual</th>
</tr>
</thead>
<tbody>
<tr>
<td>A Ab A Ab A b A b A B a B a A B a B a B a B a B a B a B a</td>
<td></td>
</tr>
</tbody>
</table>

Figure 2-1. A continuum of HL learners’ proficiency (Valdés, 2005, p. 414)
after re-acquiring the once-lost HL, whether or to what extent a HL learner can maintain the HL depends on many factors, such as personal determination, motivation, language environment, language needs, etc. Therefore, at the individual level, HL learners’ proficiency is consistently undergoing change in their lifetime (Valdés, 2005).

Since the home and community environment where HL learners grow up are quite different, the distribution of strengths in their HL also vary from individual to individual (Valdés, 2005). For instance, some HL learners are strong in receptive grammars, but limited in productive ones. They can comprehend what they are told in their HL, but are only able to respond in English. Some HL learners have no problems when engaging in an informal conversation, but cannot handle styles and registers requested in academic or professional settings. If we consider HL proficiency as a continuum, some HL learners fall into the lower extreme, while some fall into the higher extreme, and others fall in between. In other words, subgroups with various proficiencies exist in the HL learner group.

In the U.S., many postsecondary-level language programs have been trying to answer the needs of rapidly growing HL learner populations, by providing a dual-track for FL and HL learners. However, a lot of the time, HL learners are simply put in a separate track without considering the heterogeneous nature of the population. This simple dichotomy has been questioned by some researchers (Kondo-Brown, 2005; Lynch, 2008). Kondo-Brown (2005) investigated the grammar, reading, and listening skills of 185 HL learners and FL learners of Japanese. She found great similarities between FL learners, and American-born HL learners who have neither Japanese-speaking parents nor experience living in Japan. On the other hand, HL learners with at
least one native-Japanese-speaking parent presented significantly higher proficiency than the rest of the learners. Her findings indicate that HL learners with lower proficiency are more like FL learners in terms of their linguistic skills. Even within the HL learner group, the between-group difference is quite remarkable. Therefore simply separating HL learners from FL learners without considering their proficiency differences will risk putting learners with completely different language abilities in the same classroom. Lynch (2008) compared HL learners with lower-proficiency and L2 learners at advanced level. He found more similarities than differences between the two groups. I will provide more details about Lynch’s study in section 2.1.4.

2.1.3 Predictive Variables of HL Learners’ Proficiency In Their HL

HL learners come to the classroom with a wide range of linguistic knowledge and abilities, which set them apart from L2/FL learners who had no previous exposure to the target language. Even with extremely limited language input, the general sense and intuition about the language and the culture that they developed still gives HL learners advantages over their non-HL counterparts. Thus unlike non-HL learners who need to learn everything from scratch, HL learners already know something about the language. The question is: what do they know? And to what extent do they know it? Without knowing the answers to these questions, it is impossible to provide appropriate and efficient instruction to HL learners. However, finding an answer to these questions is an extremely difficult and complex task. As Hendrys (2008) argued, HL learners’ knowledge system is similar to an iceberg, which is formed by a smaller visible performance on the surface, but a much bigger invisible competence underneath. Therefore HL learners’ real proficiency is hard to assess with the current proficiency assessment methods, which are usually overwhelmingly based on performance on the
one hand, and ignore the competence on the other. Since HL learners are more likely to have a much wider range and higher level of competence compared to their performance, their proficiency levels tend to be underestimated under the current performance-oriented proficiency assessment system.

In order to access the real picture of a HL learner’s existing linguistic knowledge system, it would be best to conduct longitudinal research, observing each linguistic development stage of the learner. However, it is not realistic to do so for every learner. Nevertheless, the more we know about the background of a HL learner, the better we can infer his/her language skills. In fact, a number of social variables have been identified as possible predictors of HL learners’ proficiency.

*Learner generation* has been found as an important predictive variable of HL maintenance and attrition. Generally speaking, HL learners of an older generation usually have higher proficiency in their HL than those of a younger generation (e.g., first-generation compared to the second-generation, second-generation compared with the third-generation, etc). For first-generation immigrants, *age of arrival* is an important predictive variable: usually older arrivals with exposure to English after 12 years old show higher proficiency than younger arrivals in their HL, especially where pronunciation (Flege et al., 1995; Yeni-Komshian et al., 2000), and lexical retrieval are concerned (McElree et al., 2000). However, some individuals have lower proficiency in English. Excluding the first-generation immigrants, birth order plays an important role in predicting HL learners’ proficiency: first-born children tend to have higher levels of HL proficiency than their younger siblings (Lambert & Taylor, 1996; Zentella, 1997).
In real life, however, we sometimes encounter exceptions. For instance, a third-generation child speaks better than a second-generation child; a younger arrival is superior to an older arrival; or a third-born child speaks more fluently than the first-born child in their HL. Focusing on Spanish HL learners, Lynch tries to find explanations for these exceptions from speaker social networks. That is “HL speakers with more dense social networks including active speakers of Spanish would fare better in terms of language acquisition and maintenance than those with less dense social networks in which Spanish is rarely used” (Lynch, 2003, p. 9). In a related study, Zentella (1997), shows that the Spanish proficiency of five working-class Puerto Rican girls in New York are closely associated with the Spanish monolinguals in these people’s social networks, Thus supporting Lynch’s argument.

Based on her studies on Chinese HL learners, He argued that Cultural identity may be the most critical factor to HL development. She hypothesized that HL development “correlates positively with learner’s desire to be rooted in his/her heritage culture and to accentuate similarities” with members of the HL community (2008b, p. 116). Previous studies found HL cultural identity and HL proficiency are closely associated: Lee (2002) examined the identity and language choices of 40 second-generation Korean HL learners. The results show the stronger the identification with the Korean orientation items, the higher the Korean-language proficiency. Another study on 114 adult second-generation and first-generation (early arrival) Korean HL learners also found that those with higher Korean HL proficiency reported a stronger Korean cultural identity (Cho, 2000). Similar results were also revealed in Kondo (1997) and Tse (2000).
Socioeconomic class is another important predictive variable. Two studies on Spanish HL speakers based on Mexican-Americans in Texas revealed contradictory results. Sánchez (1983) found that those from lower socioeconomic classes tended to use more Spanish compared with those from the middle class. However, the opposite was found by Amastae (1982). Jia (2008) hypothesized that the different results found in the two studies may have been due to a generational effect, which was partially confirmed in her research on Chinese immigrants in New York. She found that lower family income significantly predicted higher self-assessed HL ability in reading, speaking, and writing. She concluded that first generation immigrants with higher income, higher educational levels, and higher English proficiency have a stronger desire to assimilate into the mainstream society. Despite the different results of the above research, socioeconomic class was proven to correlate with HL proficiency of HL learners.

2.1.4 Research on Similarities and Differences between HL and Non-HL Learners

In the U.S., the presence of HL learners in postsecondary-level language classrooms has been a challenge in the FL teaching field since the mid-1970s. The teaching professionals in this field, who were trained to teach FL learners, found themselves having to face a group of students who “in some cases, were more fluent in the language than they were, but who could not talk about the language using the terminology used in the teaching of traditional grammar,” and who have “a very difficult time learning grammar rule taught to FL students” (Valdés, 1995, p. 304). Through classroom observations, language teachers know that these learners differ from FL learners. However, one must ask: “how different are they”? Besides the differences, are there any similarities between learners of these two groups? Do methods proven to be
effective in the L2 classroom also apply to HL learners? As pointed out by Montrul, “without proper understanding of how similar or different these two types of learners are, it is difficult to tell at this point whether the exact same methods applied to L2 learners in the classroom should also be applied to heritage language learners” (2008, p. 500).

Numerous researchers investigated similarities and differences between HL and non-HL learners within the framework of UG (Isurin & Ivanova-Sullivan, 2008; Montrul, 2004; Montrul et al., 2008). They consider HL learners as incomplete learners, since the HL they “grow up speaking (or only hearing)” as an L1 was “replaced by another language as dominant and primary” (Polinsky, 2008, p. 1) at some point in their life, consequently they usually “fail to develop full linguistic ability in the heritage language” (Montrul et al., 2008). In addition, the earlier they start to become bilingual in their HL and the societal language, and “the more intense the exposure to the sociolinguistically dominant language, the more incomplete the adult grammar may turn out to be” (Montrul, 2002, p. 61). Due to the incomplete or interrupted acquisition of the HL, HL learners share some linguistic behaviors with L2 learners, such as transferring errors from the majority language. On the other hand, HL learners generally lack literacy and formal instruction in the HL, which differs from L2 learners. These UG linguists try to find out whether learners with HL background have advantages over those without HL background.

Montrul and her colleagues (2008) investigated the knowledge of Spanish gender agreement among 69 HL learners and 72 L2 learners, compared with the baseline data of 22 native speakers of Spanish. The proficiency of the HL learners was carefully controlled. Only learners who met the following three criteria were chosen: they were
born and schooled in the U.S.; had no schooling experience in their home country; and became bilingual in Spanish and English before age five. All the participants completed one oral picture description task and two written tests. The results showed that HL learners performed more like native speakers than the L2 learners on the oral test, which required more implicit knowledge; however, they were less accurate on the two written tests, which required more explicit knowledge. Moreover, like the L2 learners, the errors made by the HL learners were also systematic. Based on these results, Montrul and her colleagues argued that HL learners have “fast, implicit, and automatically processed knowledge (typically acquired early in childhood),” which developed during a critical period; but lack “meta-linguistic knowledge,” as they usually had no or very limited formal education in their HL. Therefore their knowledge of gender in Spanish “might be stored, represented, and deployed differently” from L2 learners (p. 541).

Romanova (2008) studied the morphological processing of 50 real verbs and 30 nonce verbs by 14 heritage speakers, 17 beginning L2 learners, and 28 native speakers of Russian. Of note, seven heritage speakers in her study had a few years of elementary schooling in Russia before moving to the U.S. The results showed that the heritage speakers performed closer to native speakers, and greatly outperformed L2 learners in their recognition of real verbs, especially real verbs of high-frequency. However, in terms of nonce verbs, it was the L2 learners who performed more like native speakers. Romanova attributed the differences to the limited linguistic input received by the heritage speakers in her study: “it seems that because of interrupted or incomplete first language acquisition, the weights of rules are not set and rules are
applied with a probability of less than 1" (p.119). The results also showed that heritage speakers with literacy skills consistently outperformed those without literacy skills. Romanova suggested that literacy in L1 may be a major factor influencing the development of input-based rules and probability mechanisms in morphological processing (Rothman, 2007).

A small-scale study conducted by Isurin and Ivanova-Sullivan (2008) investigated seven heritage speakers, 11 L2 learners, and five native speakers of Russian. All the participants were requested to tell a story based on a children’s picture book, *A boy, a dog, a frog and a friend*. The elicited speech samples from the three groups of learners were analyzed with a focus on selected morphosyntactic categories (e.g., aspect, case), and word order. They found that heritage speakers outperformed L2 learners in all three morphosyntactic categories.

Lynch (2008) conducted a qualitative study on five HL and four L2 learners of Spanish. The five HL learners selected in this study were either born in the U.S. or migrated here before age two, and had two to five years of formal study of Spanish. In other words, they were all typical lower-proficiency HL learners. On the other hand, the L2 learners in the study all had more than five years of formal study of Spanish, meaning they were generally at a more advanced level of Spanish L2 learning. Data were collected from individual interviews in Spanish by the researcher. Selected grammatical features were quantitatively analyzed, such as noun-adjective gender agreement, aspectual and mood distinction, subject-verb word order, etc. The overall results showed more similarities than differences between the two groups. The two
most advanced L2 learners consistently outperformed or performed equally well compared to the most advanced HL learner.

To summarize the above studies: low-proficiency HL learners, who are either simultaneous or near-simultaneous bilinguals in their HL and the societal language, show more similarities with L2 learners (Montrul, 2002). Exposure to the HL usually begins at birth or at a very early age for these learners, and comes from their family or from community members through daily life. This exposure is totally input-based, and thus these HL learners develop an implicit rather than explicit knowledge of the HL. Although their meta-linguistic knowledge may also emerge, as it does for monolingual children, there are usually no opportunities to make it fully develop through reading and writing practice, since they mostly attend English-dominant mainstream schools. With a few exceptions, most of them only develop literacy skills in English, and remain illiterate in their HL. Without formal study and literacy skills, HL learners’ linguistic knowledge of their HL is more implicit and intuitive, thus differing from traditional FL and L2 learners, who routinely acquire and store explicit linguistic knowledge of the target language through formal classroom instruction (Lynch, 2008).

Although there are so many similarities between low-proficiency HL learners and L2 learners, the exposure to the HL at an early age has lasting benefits on HL learners, even though their language experience is limited or incomplete, is interrupted by attending English-dominant mainstream schools, and undergoes attrition. These lasting benefits enable them to have advantages over L2 learners in some areas, such as phonology but not in others, such as morphosyntax (Au et al., 2002).
The differences and similarities between HL and non-HL learners provide researchers in field of SLA with a good opportunity to re-conceptualize and expand the field by “examining possible interactions between SLA and the area of language instruction currently referred to as the teaching of heritage languages” (Valdés, 2005, p. 410). Lynch (2003) argued that “the framing of a coherent agenda for research and theory building in the field of heritage language acquisition (HLA) depends in part upon the research and theory already existent in second language acquisition (SLA)…Clearly, the general sorts of questions asked in SLA are questions that HLA researchers must be asking, and the research methodologies used to respond to those questions in SLA are methodologies that would lend themselves fruitfully to HLA endeavors” (p.1). The following two studies approached HL acquisition with typical SLA research designs.

Kim (2008) investigated sentence processing differences between 88 HL and 40 non-HL learners of Korean. Based on their self-reported L1, these learners were subcategorized into five groups. HL learners were divided into Korean L1 speakers, Korean and English L1 speakers, and English L1 speakers. Non-HL learners were divided into two groups: Japanese L1 speakers, whose language is also a SOV structure like Korean, and Korean L2 learners whose L1 is a SVO structure (e.g., English, Chinese, etc). All the participants were given a picture selection task, which contained 84 test items, aiming to assess learners’ comprehension of Korean relative clauses. The results showed that HL learners with Korean as their L1 and L2 learners with Japanese as their L1 earned the highest scores among all the participants. Moreover, the analysis of error types indicated that HL learners found case markers
difficult, which affected their processing of relative clauses. Kim argued that although HL learners may not have fully acquired the syntactic structures of Korean, they can use their semantic and contextual knowledge to comprehend sentences, even when they contain complex structures. In addition, since they acquired their HL implicitly, and lack explicit meta-linguistic, or rule-based knowledge, grammar-focused instruction could play an important role in making them notice some features of the target language. Hence, “in formal language instruction, it may be crucial for teachers to provide HL learners with consistent interaction with negative feedback that draws attention to the form, rather than meaning” (p.124).

The study conducted by Gass and Lewis (2007) was the first attempt to approach the linguistic differences between HL and non-HL learners within an interactionist framework. Thirteen FL learners and six HL learners of Italian participated in the study. All the participants took part in a task-based interaction (a spot-the-difference task) with an interlocutor who was a near-native speaker of Italian. Feedback was given on non-targetlike utterances during the interaction. Learners’ perceptions about the feedback were measured using stimulated recall methodology. The results showed a between-group difference of HL and non-HL learners in terms of their perception of provided corrective feedback. Non-HL learners outperformed HL learners in perceiving phonological and morphosyntactic feedback; while HL learners perceived semantic and lexical feedback more correctly than non-HL learners. In addition, HL learners also tend to incorrectly interpret feedback on other linguistic items as semantic feedback as well. Based on these results, Gass and Lewis argued that the exposure to HL during childhood may make HL learners “regard language as a form of real communication”
whereas L2 learners who “acquire language only in a classroom setting may treat the language more as an intellectual exercise” (p.99). This study holds extremely important implications for language instruction. That is, the feedback techniques that have proven to be effective for FL/L2 learners may also work for HL learners, but in a different way.

2.2 CHL Learners

2.2.1 Chinese Dialects

Chinese is an umbrella term for seven major dialects in the Chinese language family: Mandarin 北方话, Wu 吴, Xiang 湘, Gan 赣, Min 闽, Cantonese 广东, and Hakka 客家 (He, 2006). These seven dialects can be subdivided into two groups based on geography: Northern Chinese and Southern Chinese. Mandarin is a label applied to dialects spoken in Northern China. More precisely, it refers to dialects spoken along the Yellow Plain and the Loess Plateau, including Northwestern, Northern proper, River, and Southwestern varieties. Despite having these varieties, Mandarin is highly unified. Different dialects of Mandarin are mutually intelligible. On the other hand, Southern Chinese consists of six mutually unintelligible dialects: Wu 吴, Xiang 湘, Gan 赣, Min 闽, Cantonese 广东, and Hakka 客家. The biggest obstacle for Northern Chinese and Southern Chinese speakers to understand each other is the phonological differences among the dialects: generally speaking, Southern Chinese dialects tend to have more tones than Northern Chinese dialects. In addition, Southern Chinese dialects also have sounds which do not exist in Northern Chinese, such as stop endings (Sun, 2006).

In the SLA literature, Mandarin refers to the standard dialect of Chinese officially recognized by the governments in mainland China, Taiwan, Hong Kong, and Singapore. It is based on the Beijing dialect phonological system as its “norm of pronunciation, and
Northern dialects as its base dialects” (Chen, 1999, p. 24). It is called putonghua 普通话 “common speech (language)” in mainland China and Hong Kong, guoyu 国语 “national language” in Taiwan, and huayu 华语 “Chinese” in Singapore. It is also the widely taught variety in Chinese language classrooms (He, 2008a). Henceforth, the term Mandarin only refers to the standard dialect of Chinese in this dissertation.

In mainland China, the government has played an active role in promoting Mandarin, or putonghua 普通话, since the 1950s. According to survey results released by the National Language Committee (2005) 53% of the Chinese population in mainland China speaks Mandarin. The percentage of the Mandarin-speaking population varies dramatically with age and education. Generally speaking, the younger the age, and the higher the education, the higher the percentage of Mandarin speakers within the population. For instance, while 40.59% of people from 45-59 years old speak Mandarin, the percentage jumps to 70.12 % when the age of the population decreases to between 15 and 29 years old. In terms of education, while 25.49 % of people with elementary-school education speak Mandarin, 86.77% of people with more than two-years of college education are Mandarin speakers.

Compared with mainland China, Taiwan was more strict in promoting Mandarin, or guoyu 国语, in the 1980s, at the expense of other Chinese dialects and aboriginal languages. All dialects other than guoyu 国语 were banned in schools and mass media. As a result, Taiwan, which regarded guoyu 国语 as “the High language and the local Chinese dialects and aboriginal languages as the low languages”, was more “characterized by bilingualism and diglossia” than mainland China (Chen, 1999, p. 60). In 1991, about 90% of the population of Taiwan spoke guoyu 国语, which was much
higher than the percentage of the Mandarin-speaking population in mainland China at the same era. At the same time, 82.5% of native speakers of Southern Min—the most widely spoken dialect other than Mandarin in Taiwan—are bilingual in Southern Min and "guoyu 国语"; 88% of native speakers of other Chinese dialects and aboriginal languages are bilingual in "guoyu 国语" and their native tongue. However, since the late 1980s, there has been an increased awareness and enhancement of regional identity and local dialects. In 1987, the Ministry of Education formally allowed school children to speak dialects at school. Consequently, the use of Southern Min has dramatically increased in the society (Huang, 1993).

Although 97% of Hong Kong people speak Cantonese as their native language, English instead of Cantonese or Mandarin had been regarded as the High language until July 1997, the year when Hong Kong became a Special Administrative Region of China. However, after 1997, the use of Cantonese in public functions of the government has been dramatically increased. "Putonghua 普通话" is also becoming more and more important in the society (Chen, 1999).

2.2.2 Chinese People in the U.S.

According to the U.S. Census Bureau, in 2000, among the over 2.2 million people of Chinese ethnicity in the United States, 83% of them spoke Chinese, while the rest of them were either English monolinguals or speakers of other languages. Moreover, the majority of the Chinese speakers were foreign-born, and ranked as the fourth-largest immigrant group in the U.S.

These people, or their ancestors, moved to the U.S. mainly as the result of three waves of large-scale immigration from China to the U.S. in the past 160 years. The first
wave can be traced back to 1849, when the California gold rush occurred. At that time, Chinese immigrants were poor farmers with no literacy skills, who were mainly from Guangdong Province and spoke solely Cantonese. The second wave of Chinese immigration to the U.S. took place between the 1950s and 1970s. The Chinese who immigrated during this period usually came from more educated and wealthy backgrounds than their Cantonese-speaking predecessors. Among them, those from mainland China and Taiwan spoke mostly Mandarin, while those from Hong Kong spoke Cantonese. The largest and still ongoing Chinese immigration started in the 1980s, and has continued to date. The majority of these immigrants are scholars from mainland China. Regardless of which province in China they are originate from, in addition to the local dialects (if there are any), these people all speak Mandarin (Chang, 2003).

The above brief historical sketch shows that the language diversity that currently exists among the Chinese in the U.S. correlated with the historical fact that these people or their ancestors originally came from different parts of China where various dialects were spoken. Based on this observation, people who migrated from mainland China and Taiwan speak better Mandarin than those who were from Hong Kong. This observation was partially proven by an on-line survey conducted in 2006, by Wiley and his colleagues. They investigated Mandarin and other dialects spoken among Chinese immigrants and international students in the U.S. Overall, more than 50% of the respondents claimed that they spoke Mandarin well or very well regardless of the regions they were originally from. However, people from mainland China were most confident about their fluency in Mandarin. In fact, 91.3% of them reported that they
spoke Mandarin very well, which was higher than the percentage reported by people from Taiwan (79.8%). Only 18.4% of people from Hong Kong thought they spoke Mandarin very well (Wiley et al., 2008). Although international students and HL speakers were not differentiated in the survey, the results still show that the regions of origin of the Chinese people in the U.S. can be used as a useful variable to predict their proficiency in Mandarin.

2.2.3 Defining CHL Learners

As with HL learners in general, CHL learners are defined either broadly, based on their ethnic, historical, and family connection to the language, or narrowly, according to their actual linguistic proficiency in Chinese.

Following Fishman (2001), CHL learners can be defined as individuals who have a personal interest or involvement in Chinese, a language spoken by their ancestors (Li & Duff, 2008). This definition includes a wide range of learners with various language backgrounds, such as learners who are ethnically Chinese but with no or limited exposure to Chinese language and culture, learners who speak Chinese fluently but with no literacy skills, learners who speak a dialect other than Mandarin, etc.

Based on actual linguistic competence and familial affiliation, following Valdés (2001), a CHL learner is defined as a learner who “is raised in a home where Chinese is spoken and who speaks or at least understands the language and is to some degree bilingual in Chinese and English” (He, 2006, p.1).

However, CHL learners identified under this definition still have a very “uneven grasp of the HL, falling along a continuum of having very little HL knowledge to being highly proficient” (Li & Duff, 2008, p. 17). Like HL learners of other languages, the cause of CHL learners’ uneven proficiency distributions relates to their different language
experiences; however, the following two language-specific factors also play important roles.

The first is the home dialect factor. As Valdés (2001) pointed out, in many cases immigrant students are speakers of non-prestige varieties of their HL. In terms of CHL learners, as we have seen above, they or their ancestors were originally from different hometowns where various Chinese dialects are spoken. Therefore their home dialects also vary accordingly. For instance, HL learners who are originally from Hong Kong mostly speak Cantonese at home, while those from mainland China more likely speak Mandarin, or a mix of Mandarin and other dialect(s) (Wiley et al., 2008).

In the U.S., Mandarin is the most widely taught variety of Chinese in Chinese language classrooms. According to a survey conducted by the Modern Language Association (MLA), in 2006, 51,582 students enrolled in Mandarin class at the postsecondary-level, which ranked as the 7th most commonly taught language other than English in the U.S. On the other hand, other dialects of Chinese are hardly even introduced into mainstream educational institutions. According to the same survey, the enrollments in Cantonese classes were only 178, while only 21 students took Taiwanese, which is one of the varieties of Min spoken in Taiwan. No records were available for other dialects. This indicates that those dialects are either not taught in the classroom at all, or the enrollments are too small to be considered in the survey (Modern Language Association, 2006).

Thus for many HL learners in the U.S., learning their HL in a postsecondary-level classroom means learning a second dialect (Valdés, 2005) when their home dialect is
not Mandarin. This is more challenging for learners whose home dialect is not mutually intelligible with Mandarin, such as Cantonese, Min, etc.

The second factor is the orthography system. Chinese people all use one unifying writing system. Unlike English and other phonographic writing systems, the Chinese writing system is a logographic system with “each grapheme (or character) simultaneously encoding sounds and meaning at the level of the syllable…representing various concepts or ideas totally divorced from pronunciation” (Sun, 2006, p.102). Therefore, regardless of which dialect is spoken, a literate Chinese person can read any character with his/her own local pronunciation, and write sentences that can be understood by people from any dialect without being able to speak that dialect. Thus, unlike English, literacy in Chinese is measured by the number of characters a learner has mastered rather than the number of words. In mainland China, a student is required to master minimally 2,500 characters when graduating from elementary school, and 3,500 characters when graduating from college (Chen, 1999).

It was estimated that there are over 56,000 Chinese characters, with 3,000 of them commonly used in modern Chinese. As learning characters is so difficult and time-consuming for Chinese children, the simplification of Chinese characters started in the early 1900s. In mainland China, between the 1950s and the mid-1960s, over 2,000 of the most commonly used characters were officially simplified, and used in mainland China. This is known as \textit{jianhuazi} 简化字 “simplified Chinese”. Today simplified Chinese is also officially used in Singapore (He, 2008a). In Taiwan and Hong Kong, however, the older script, which is known as \textit{fantizi} 繁体字 “traditional Chinese”, is still used. Therefore in the U.S., HL learners, whose families are originally from Taiwan and Hong
Kong tend to support instruction in traditional script, while those from mainland China prefer simplified script (Chen, 1999; Sun, 2006). Thus there is a high likelihood that HL learners, even with some literacy skills, have to re-learn the writing systems when their home script is different from the school script.

In addition, regardless of the simplified or traditional script, there exists a much greater distance between spoken Mandarin and its orthography than that in a language with a phonographic writing system. Therefore, the unbalance and disparity in the four language skills tends to be even more pronounced in CHL learners than HL learners of languages with a phonographic writing system. CHL learners are, by definition, bilingual, but are not very often bi-literate, when they arrive in our classrooms (Dai & Zhang, 2008; Hendryx, 2008; Koda et al., 2008b).

To summarize, when considering the proficiency of CHL learners in the U.S., besides the factors which are observed in HL learners of other languages, we also need to look at learners’ home dialects and writing scripts. Thus a learner from a Mandarin speaking family may still have to face the situation that the classroom script differs from his/ her home script. A learner whose home dialect is unintelligible to Mandarin may have difficulty with aspects of the spoken language, but may not have equal trouble in writing if the classroom script is the same as his/ her home script (He, 2008a). Many Cantonese speaking learners belong to this category.

Based on classroom observations at the postsecondary level in Chinese programs, Hendryx identified five commonly-seen subgroups among CHL learners (Hendryx, 2008).
Learners in the first group usually have little command of the HL language, but do not lack intuitions about it. In addition, they also store a considerable amount of cultural knowledge. Learners in the second group have slightly better abilities than those in the first group, especially in terms of their speaking and listening skills. However, these skills are usually more receptive rather than productive. They usually have no or quite limited abilities in reading and writing. To summarize, CHL learners in the first and second groups fall into the group of lower-proficient HL learners (Lynch, 2008), who share more similarities with L2 learners in terms of their language skills and linguistic behavior (Kondo-Brown, 2005; Li, 2008; Lynch, 2003). Nevertheless, their limited exposure to the HL still gives them some advantages over those without any HL experience, which enables them to acquire the target language faster than L2 learners.

Learners in the third group have excellent speaking and listening skills, but very limited or even no reading and writing skills. On the contrary, learners in the fourth group have fairly developed reading and writing skills, but little abilities in Spoken Mandarin. In the U.S., most of these latter learners are from Cantonese, and Taiwanese speaking families. For them, HL learning means learning the second dialect.

Learners in the fifth group have near-native proficiency in all four language skills. These learners all received formal or semi-formal education in the target language, either through after-school programs, community-based schools, or by studying at schools in their home country.

2.2.4 The Differences between CHL and Non-CHL Learners

In the U.S., CHL instruction and studies is a newly-emerging field with approximately 10 years of history. Since the mid 1990s, it has attracted a rapidly growing body of scholars from various disciplines, such as SLA, bilingualism, reading
research, discourse analysis, orthography analysis, and/or language pedagogy (He, 2008a). In these scholars’ studies, CHL learners presented unique characteristics, which differ from their non-CHL counterparts in terms of linguistic knowledge system (Dai & Zhang, 2008; Hendryx, 2008), literacy skills (Jia, 2009; Koda et al., 2008b; Xiao, 2008), morphological awareness (Koda et al., 2008a), Chinese character learning (Ke, 1998; Xiao, 2006), and their motivation (Lu & Li, 2008; Weger-Guntharp, 2006), and identity (He, 2004; He, 2008b).

**Linguistic proficiency.** CHL learners tend to have uneven skills in speaking, listening, reading, and writing (Hendryx, 2008; Weger-Guntharp, 2006; Wu, 2008), which result from their heterogeneous past exposure to and learning experience in Chinese. Their CHL acquisition usually starts at home, relying mainly on a vertical and reciprocal intimate relation with their parents or grandparents. With very few exceptions, their acquisition heavily emphasizes speaking and listening rather than reading and writing skills (Dai & Zhang, 2008). Consequently, most of them are bilingual in English and Chinese, but not bi-literate, particularly when little or limited print materials are available at home (Koda et al., 2008b). In addition, because of the dominance of English language and limited exposure to the HL, HL learners usually experience great decline of their HL skills right after they start schooling in mainstream schools, which is known as language attrition (Fishman, 1991; Wong Fillmore, 1991). R. Jia (2008) investigated 85 Chinese HL speakers in New York City, aged from 16 to 30 years old. More than 90% of them were foreign-born and emigrated to the U.S. before 20 years old. Data used in this study were collected through a tailor-made language background questionnaire, which asked learners to self-rate their proficiency in both languages for
each two-year segment of the entire time span from their systematic exposure to English. The results showed that the participants’ HL skills all experienced great attrition after five years of residence in the U.S. on the one hand, and a great improvement in English on the other hand. Moreover, among the four skills, reading and writing abilities showed larger scales of attrition compared with speaking and listening skills. In contrast, non-CHL learners usually have relatively even distribution of the four language skills, which almost entirely developed through formal classroom learning.

Similar results were found by Xiao (2006), who conducted two studies comparing the language skills of the two learner groups. The first one compared 18 CHL and 18 non-CHL learners from a high-beginning-level intensive Chinese course in terms of their listening, grammar, reading, and character writing skills. The results showed that CHL learners significantly outperformed non-CHL learners in listening and grammar, but not in reading comprehension and character writing. The second study was conducted among 94 HL and 54 non-HL learners of Chinese from three different instructional levels: beginning, intermediate, and advanced. A 25-item grammaticality judgment test and a 6-item English-to-Chinese translation task were distributed during a 20-minute class time. The results showed that HL learners did significantly better on the grammaticality judgment test, but not on the translation task.

Ke (1998) also found that CHL learners had no significant advantages over non-CHL learners in terms of Chinese character learning. He conducted a study among 60 FL and 85 HL learners of Chinese, who just finished their first year of Chinese instruction at seven institutions. The demographic background information of the HL participants in this study is quite limited, and only briefly mentioned that these
participants “had been exposed to some form of Chinese orally and aurally at their homes” (p.94). No information is available regarding the participants’ previous HL literacy experience (e.g., home schooling, attending community schools, etc.). All the participants were given one character recognition task, and one character production task. Each test consists of 30 characters chosen from the first-year Chinese textbook used by these learners. The results showed no significant difference between the two groups of learners in terms of their recognition and production of Chinese characters.

In sum, CHL learners are superior in listening, speaking, and grammar, which could be attributed to the large amount of HL input they are exposed to at home through daily oral communication with their family members, even though their skills are more receptive rather than productive. This advantage is not shared by their non-CHL counterparts. In contrast, literacy skill development requires more systematic learning and rich HL print resources at home (Xiao, 2008). Compared with English, which has inflectional morphemes in structurally transparent alphabet-based words, the Chinese writing system is a logographic system, which does not phonologically associate with the character. Therefore, character learning is considered the hardest part in Chinese learning (Everson, 1998). Many Chinese parents try to teach their children at home. However, whether the learning could proceed in a systematic manner depends on parents’ persistence, time, and educational background, etc. In many cases, parents give up trying (Hinton, 1999), and send their children to community weekend Chinese schools for their CHL literacy development. However, children usually lose their CHL literacy skills after formal schooling since their home literacy does not relate to their school literacy (Wong Fillmore, 1991). Thus many CHL learners are illiterate or with
very limited literacy skills even though they are fluent in the oral language (Li, 2006). As a result, for CHL learners, the goals for their HL learning are more likely to improve their literacy skills in order to read and write well. In contrast, for non-CHL learners, they rank speaking and listening as the most important immediate goals in their Chinese learning (Wu, 2008).

**Cultural identity.** CHL learners bring to the classroom not only linguistic skills, but also their cultural identity, which has been influenced by both Chinese and American cultures. This is also not shared by their non-CHL counterparts.

Similar to their linguistic skills, their cultural identities also vary from individual to individual, depending on the home and environment they grew up in. Dai and Zhang (2008) conducted a survey among 80 college students with CHL background, and their participants perceived their own cultural identities differently. The majority of the participants viewed themselves as a combination of both Chinese and American cultures, and they drew on different cultural knowledge when interacting with different people, at different times, and in different social contexts. The rest of the participants either considered themselves “less Chinese than the Chinese, and less American than the Americans” (p.44), and thus indicated that they belonged to neither Chinese or American culture, or accepted one culture but rejected the other. G. Jia (2008) found that HL speakers who identified a stronger connection and preference to their HL culture tended to not only use more HL, but also self-rate their reading and writing skills higher.

In a related study, He considers learners’ identity as a prime dynamic force rather than just the background in CHL learning. She argued that through learning CHL, which
takes place in a three-dimensional framework with intersecting planes of time, space, and identity, a CHL learner will form a new cultural identity which inherits “some of the ‘Chineseness’ from his family and his neighborhood but will enable him to become a very different kind of Chinese-American from his family and his neighbors”. (He, 2008b, p. 110). In addition, she notes that, “the degree to which a learner’s CHL develop is dependent upon the degree to which s/he is able to find continuity and coherence in multiple communicative and social worlds in time and space and to develop hybrid, situated identities, and stances” (p. 116). In her identity-based model for CHL development, she hypothesized that the degree of success in CHL development positively correlates learners’ desire to be connected with their heritage culture and CHL community members for the long run, the desire to communicate in CHL in a moment-by-moment manner, the home language of the learner, the availability and diversity of the CHL input, as well as the extent to which the learner has created a niche in the English-speaking linguistic, social, and cultural community.

**Motivation.** Chinese was ranked as one of the most difficult languages for native English speakers to learn (Hadley, 2001). Therefore regardless of CHL or non-CHL learners, a strong motivation must be present, since it is one of the key factors influencing the rate and success of second/foreign language (L2) learning (Dörnyei, 1994; Ely, 1986; Gardner, 1985). Gardner defined motivation as “the extent to which the individual works or strives to learn the language because of a desire to do so and the satisfaction experienced in this activity” (Gardner, 1985, p. 10). Gardner’s well-known socio-educational model distinguished two types of motivations: instrumental and integrative. The former one refers to a pragmatic reason that drives learners to learn a
FL or L2, such as career development, travelling, passing exams, moving to a new country, etc. On the other hand, integrative motivation closely relates to a desire of getting closer or even becoming the member of a language community. Lu and Li (2008) conducted a comparative analysis of the effect of different motivational factors on 59 CHL and 61 non-CHL learners from nine Chinese college classes. All the participants were required to complete a three-part questionnaire, consisting of 63 questions, followed by an interview. The results showed that although both types of motivation played an important role in the Chinese learning by participants from both learner groups, CHL learners were significantly more influenced by instrumental motivation than their non-CHL counterparts. In addition, their integrative motivation was more highly correlated to their perceptions on their listening and speaking abilities, but not their reading or writing skills. For the non-CHL learners, the correlation was found between their integrative motivation and listening, speaking, as well as writing, but not reading skills.

In sum, the previous research revealed differences between CHL and non-CHL learners in linguistic, cultural and social aspects. Although research in this field has made great progress in recent years, the majority of these studies were non-experimental, the data in which were mainly collected through questionnaires, interviews, classroom observations, case studies. To further explore the cause-and-effect questions on the learning process and developmental path of CHL learners, experimental studies are certainly necessary. To my knowledge, no experimental research has been conducted in comparison of the two learner groups under an
interactionist framework, such as the benefits of different types of corrective feedback, which motivated the current study.

1 In the current study, the term *dialect* refers to all the local varieties spoken in mainland China, Taiwan, Hongkong which share a high degree of uniformity in terms of grammar and writing system.

2 All the Chinese characters used in this paper take the form of simplified Chinese.

3 *Unintelligible* here refers mainly to the phonological differences between Southern and Northern dialects. Both the two dialects share the uniformed writing system. In the current study, speakers of both Southern and Northern dialects were also considered as native speakers of Chinese.
CHAPTER 3
IMPLICIT AND EXPLICIT FEEDBACK

This chapter will focus on implicit and explicit feedback. First, two pairs of related theoretical constructs—implicit vs. explicit learning and implicit vs. explicit knowledge—will be introduced, followed by the definition of implicit vs. explicit feedback. The next section will review the main measures that have been employed in the literature in examining the effects of corrective feedback. Then the discussion will move to the effects of feedback on L2 learners’ language development based on a review of 18 previous studies. Lastly, implicit and explicit feedback will be discussed in relation to HL learners, who are the target learner population in the current study and have been less studied in the past.

3.1 Theoretical Constructs

3.1.1 Implicit vs. Explicit Learning

Learning a language can be either implicit or explicit. Young children learn their L1 in a manner as effortless as they grow. That is they not only do not have any goals or motivations for learning, but also do not even realize they are learning. It seems that as long as you immerse children in the target language environment, they can automatically become a competent speaker in that language. No explicit instruction or correction is needed (Rober, 1989). In this sense, children’s L1 acquisition is unconscious rather than conscious. In terms of complex grammar, children simply rely on their exposure to whatever linguistic data is available around them in a natural setting, such as communication with adults or older children, TV programs, etc. Linguists working from an innate position claimed that children are born with a set of abstract principles and parameters that form the underlying grammars of all natural
languages. Therefore they can completely master the language with incomplete input in the language data, regardless of whether it is well-formed or not (Chomsky, 2001). In contrast, linguists working from an interactionist position disagree that UG is the only determining factor for children’s language; though they admit that children are born with some sort of innately specific knowledge, they argue that it interacts with the environment and input (Gass, 1997).

Despite these discrepancies, linguists generally agree that children’s L1 acquisition is different from adults’ L2 acquisition. What works perfectly for children may not work equally well for adults. Compared to children, adults’ L2 acquisition is goal-oriented, and motivated by various reasons. They usually need to acquire rules in order to master grammar (at least certain features of grammar), instead of simply relying on unconscious exposure to natural linguistic data (R. Ellis, 1994). Thus learning a L2 can occur through either implicit learning or explicit learning. Implicit learning refers to “acquisition of knowledge about the underlying structure of a complex stimulus environment by a process that takes place naturally, simply and without conscious operations,” whereas explicit learning is “a more conscious operation where the individual makes and tests hypotheses in search for a structure” (R. Ellis, 1994, p. 1). The fundamental difference that sets these two types of learning apart is whether a learner is conscious of what is being learned (Dekeyser, 1998).

Conscious here includes at least two meanings: the expressibility issue and the intuitive judgment issue. Expressibility is how well a learner can verbally express the acquired information. For instance, children cannot verbalize why they say ‘he goes to school everyday’, but ‘he went to school yesterday’. So they have only procedural
knowledge but not declarative knowledge, both of which are stored in their long-term memory (Chamot & O'Malley, 1994). Thus their linguistic knowledge is more implicit than explicit. *Intuitive judgment* refers to the notion that the acquired information is “presumably not present in conscious awareness and not used deductively as the basis for making grammaticality judgments in the testing phase” (Rober, 1989, p. 230). In other words, when making judgments, learners do not bring the knowledge resulting from implicit learning to the awareness level, instead the judgments are intuitive.

Krashen (1981; 1982; 1985) views implicit and explicit learning as two types of learning with contrastive features. He clearly distinguished language *acquisition* from language *learning*: *acquisition* is incidental and implicit, with no selective attention to features of input that feed into the learning process, which results in *knowing* the language; *learning*, on the other hand, is intentional and explicit, with selective attention, which results in *knowing about* the language. Krashen argued that only implicitly acquired knowledge can be the basis of fluent production, which is particularly effective for complex structures. He claimed that the traditional explicit grammar instruction and error correction in second language classes had only very limited effects on learners’ language development.

Many researchers did not agree with Krashen, particularly about the role of explicit learning. They argued that explicit learning is effective in L2 acquisition, which was supported by empirical evidence from laboratory studies (e.g., DeKeyser, 1995; Leow, 1998). Some of them believe that explicit learning plays a significant role in L2 acquisition, because it could help learners notice the gap between input and their own output (R. Ellis, 1994) through the explicit teaching of grammar, explicit error correction,
or input enhancement (Norris & Ortega, 2000). However, they think the role of practice is only limited to item learning and fluency improvement. Some researchers believe that explicit learning works effectively in L2 acquisition through systematic practice (e.g., DeKeyser, 2003; Schmidt, 1990, 1994, 1995; Schmidt & Frota, 1986; Swain, 1985), because practice could “gradually bridge the gap between explicit knowledge and language use” (DeKeyser, 2003, p. 328).

3.1.2 Implicit vs. Explicit Knowledge

There are two types of linguistic knowledge internalized in learners' mind: implicit knowledge and explicit knowledge. Implicit knowledge refers to “knowledge that learners are only intuitively aware of and that is easily accessible through automatic processing”, whereas explicit knowledge consists of “knowledge that learners are consciously aware of and that is typically only available through controlled processing” (R. Ellis et al., 2006, p. 340). Implicit knowledge is *knowledge of language*, which can be further distinguished into two types: formulaic knowledge, which is internalized in learners’ mind as unanalyzed units. For instance, NS usually know a large number of fixed or semi-fixed expressions. Another type is rule-based knowledge, which facilitates learners to produce novel sentences. On the other hand, explicit knowledge is *knowledge about the language*, which is analyzed, abstract, and explanatory (R. Ellis, 1994). Learners usually are able to verbalize explicit knowledge if called upon. However, this knowledge is not dependent on meta-linguistic knowledge (Bialystok, 1994, p.567).

Implicit and explicit knowledge distinguished from each other mainly in the following areas: awareness, type of knowledge, systematicity and certainty of L2 knowledge, accessibility of knowledge, use of L2 knowledge, self-report, and learnability (R. Ellis, 2005), which are summarized in Table 3-1. Implicit knowledge is intuitive, in the sense
that “the learner is unlikely to be aware of having ever learnt it and is probably unaware of its existence” (R. Ellis, 1994, p. 85). On the other hand, explicit knowledge is “analyzed (in the sense that it exists independently of the actual instances of its use), abstract (in the sense that it takes the form of some underlying generalization of actual linguistic behavior) and explanatory (in that the logical basis of the knowledge is understood independently of its application)” (R. Ellis, 1994, p. 85). Implicit knowledge involves automatic processing. It is usually not verbalizable, and some argue that certain aspects of it can never be learned after a certain age. On the other hand, explicit knowledge can be viewed as “a tool that learners use to achieve control in demanding situations”, and that involves controlled processing. It is potentially verbalizable, and can be learned at any age (R. Ellis, 2005, p.150).

**Table 3-1. Key characteristics of implicit and explicit knowledge (R. Ellis, 2005, p.151)**

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Implicit knowledge</th>
<th>Explicit knowledge</th>
</tr>
</thead>
<tbody>
<tr>
<td>Awareness</td>
<td>Intuitive awareness of linguistic norms</td>
<td>Conscious awareness of linguistic norms</td>
</tr>
<tr>
<td>Type of knowledge</td>
<td>Procedural knowledge of rules and fragments</td>
<td>Declarative knowledge of grammatical rules and fragments</td>
</tr>
<tr>
<td>Systematicity</td>
<td>Variable but systematic knowledge</td>
<td>Anomalous and inconsistent knowledge</td>
</tr>
<tr>
<td>Accessibility</td>
<td>Access to knowledge by means of automatic processing</td>
<td>Access to knowledge by means of controlled processing</td>
</tr>
<tr>
<td>Use of L2 knowledge</td>
<td>Access to knowledge during fluent performance</td>
<td>Access to knowledge during planning difficulty</td>
</tr>
<tr>
<td>Self-report</td>
<td>Nonverbalizable</td>
<td>Verbalizable</td>
</tr>
<tr>
<td>Learnability</td>
<td>Potentially only within critical period</td>
<td>Any age</td>
</tr>
</tbody>
</table>

Although linguists generally recognize the important role played by implicit knowledge in L2 acquisition, they dispute about the role of explicit knowledge. In
addition, the relationship between the two types of knowledge has also been much debated. These controversies resulted in very different pedagogical recommendations. There are three major positions: the non-interface position, the weak-interface position, and the strong interface position.

Krashen’ (1981) non-interface model completely denies any associations between implicit and explicit knowledge. He believes that these two types of knowledge have their independent mental processes and storage. Explicit knowledge can never be converted into implicit knowledge, and vice-versa. He claims that implicit and explicit knowledge are caused by different acquisitional mechanisms. Implicit knowledge is the result of acquisition, whereas explicit knowledge is caused by learning. Learned competence does not become acquired competence. The results of learning can never lead to implicit knowledge. Therefore the role of L2 instruction should really be to provide large quantities of comprehensible input to foster implicit learning, not to provide explicit rules and systematic practice of these rules. The acquired implicit knowledge is used by learners in producing utterances; whereas the learned explicit knowledge only functions as a monitor to edit the utterances produced by acquired knowledge, which is known as the Monitor Hypothesis (Krashen, 1981; 1985). Krashen’s position was criticized for not operationable as it failed to provide precise definitions of the two critical constructs in this model: acquisition and learning (e.g., DeKeyser, 1997; Gregg, 1984).

In contrast, DeKeyser (1998) in the strong interface position claims that explicit knowledge can not only be derived from implicit knowledge, but also be converted into implicit knowledge through intentional practice. He claims that L2 knowledge is learned by learners as declarative knowledge first, and then changed into procedural knowledge
through communicative practice. Declarative knowledge is defined by DeKeyser as "factual knowledge", which can be slowly restructured as procedural knowledge through repeated target behaviors. On the other hand, procedural knowledge is defined as behavior involving "condition-action pairs that state what is to be done under certain conditions" (p.48). The strong interface position recommended that grammar rule should be taught and then practiced in order to gradually "bridge the gap between explicit knowledge and use" (DeKeyser, 2003, p. 328), such as let learners practice in communicative tasks. In the weak-interface model, which was developed by R. Ellis (1993; 1994), explicit knowledge is considered to play an important role in L2 acquisition, particularly facilities learners to notice some specific linguistic features of the input through comparing their own output and the input. This position allows two-way conversions between implicit and explicit knowledge. However, these conversions are only allowed under certain stringent conditions: explicit knowledge can be converted into implicit knowledge when learners are developmentally ready to acquire a new form; implicit knowledge can also be converted into explicit knowledge when explicit knowledge can be used to monitor learners' output from implicit knowledge. R. Ellis's proposes that explicit and implicit knowledge can be taught separately: explicit knowledge can be taught through consciousness-raising tasks, whereas implicit knowledge can be taught in task-based activities.

3.1.3 Implicit vs. Explicit Feedback

In SLA, negative feedback is a mechanism that responds to errors made by learners during their practice. It provides information regarding the accuracy of learners’ utterances (Leeman & Martinez, 2007). Negative feedback can be either implicit or explicit, depending on how an error was indicated: when the error was covertly
indicated, the feedback is considered implicit; when the error was overtly indicated, it is considered explicit (Ellis et al., 2006).

In the previous studies, the most prevalent form that has been taken for implicit feedback is recasts (e.g., Ammar & Spada, 2006; Carroll, 2001; Carroll & Swain, 1993; R. Ellis, 2007; Iwashita & Liem, 2005; Kang, 2009; Leeman, 2003; Long, 1996; Lyster, 2004; Muranoi, 2000; Nassaji, 2007; Sheen, 2007). Besides recasts, implicit feedback was also operationalized in the forms of clarification requests (e.g., Loewen & Nabei, 2007; McDonough, 2005). On the other hand, explicit feedback has been operationalized in a variety of forms. Some of the studies involve only minimal explicit feedback, such as explicit feedback in an implicit condition (Rosa & Leow, 2004), negative evidence (Leeman, 2003), repetition (e.g., Kang, 2009; Lyster, 2004), or elicitation of the correct form (e.g., Lyster, 2004; Ammar & Spada, 2009). Others involve more detailed explanation, such as explicit rejection (e.g., Carroll & Swain, 1993; Kim, 2008), meta-linguistic feedback (e.g., Ammar & Spada, 2006; R. Ellis et al., 2006; Kim & Mathes, 2001; Loewen & Nabei, 2007; Lyster, 2004), and explicit correction (DeKeyser, 1993).

Recasts generally refer to a NS’s reformulation of all or part of a NNS’s utterance, minus the error (Lyster & Ranta, 1997). Long provided a more detailed definition, in which a recast is defined as “a reformulation of all or part of a learner’s immediately preceding utterance in which one or more nontarget-like (lexical, grammatical, etc.) item(s) is/are replaced by the corresponding target language form(s), and where, throughout the exchange, the focus of the interlocutors is on meaning, not language as object” (Long, 2007, p.77). Long’s definition highlighted several important characteristics
of recasts, which set them apart from other types of negative feedback: first, the corrections in recasts are done in an implicit and incidental manner, with no clear indicator of the source of errors. Therefore, learners have to infer by themselves that their utterances are problematic (Carroll & Swain, 1993). Second, the juxtaposition of the learner’s ungrammatical utterance and the teacher’s reformulation may enhance the salience of the target form and promote learner noticing, which provides the learner with an ideal opportunity to make a cognitive comparison and to notice the gap between the target-like and non-targetlike forms (Doughty & Pica, 1986; Long et al., 1998). Third, since corrections are embedded in a context, the flow of communication will not be interrupted (Long, 1996, 2007). As a result, recasts can serve two functions: a pragmatic function that keeps the communication channel open, and a corrective function that indicates that an error was made by the learner. Thus the function of recasts can often be ambiguous, as it can be used both for providing and seeking confirmation and for additional information (Lyster, 1998b). In fact, classroom studies showed that corrective recasts were more likely to be interpreted by learners as non-corrective repetitions during the discourse in SLA classrooms, as both of these two types of feedback share the same pragmatic functions (Lyster,1998a).

Clarification requests can be used for problems in either comprehensibility or accuracy, and include phrases such as ‘pardon me’, ‘what do you mean by X’? (Lyster & Ranta, 1997). Similar to recasts, the corrective intention of clarification requests is also not so easily noticed because of the dual purpose it serves: clarifying meaning as well as prompting learners to self-repair their problematic utterances. Therefore, it is also considered more implicit than explicit (Loewen & Nabei, 2007).
Explicit feedback in an implicit condition was operationalized by Rosa and Leow (2004) as a simple indication of whether the answers made by learners are correct or not. Negative evidence, which was employed in Leeman’s study, was operationalized as feedback containing evidence “on ungrammatical noun-adjective agreement” (2003, p. 49). With repetition, teachers “replicate the student’s error verbatim, usually with rising intonation and stress to highlight the error” (Lyster, 2004, p. 405). Elicitation, in contrast, can include three techniques: first, using direct questions such as ‘how do we say that in French?’; second, pausing to allow the student to complete the utterance initiated by the teacher; third, asking the learner to reformulate the utterance (Lyster & Ranta, 1997).

In Carroll and Swain’s study (1993, 2001), learners in an explicit hypothesis rejection group were told that they made an error, followed by an explicit explanation of that error. Meta-linguistic feedback has been frequently compared with recasts regarding the impact of each on L2 development (e.g., Ammar & Spada, 2006; Carroll & Swain, 1993; R. Ellis et al., 2006; Kang, 2009; Kim & Mathes, 2001; Sauro, 2009; Sheen, 2007). The former usually contains “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). As the correct linguistic form is not provided in the meta-linguistic feedback, learners are pushed to produce the correct form by themselves based on the provided meta-linguistic information.

The above dichotomy of implicit and explicit feedback has been questioned by researchers (e.g., R. Ellis et al., 2006; Loewen and Nabei, 2007), who argued that the line between these two types of feedback is not always clear: to what extent should
feedback be overt enough to be considered explicit? How covert must feedback be to be categorized as implicit? For instance, although recasts are generally considered implicit, not all of them are as implicit as is generally assumed. In fact some of them are quite explicit, especially when saliency cues (e.g., phonological emphasis) are involved (R. Ellis et al., 2006). These cues refer to “additional intonational or verbal cues or signals to highlight the error and/ or push the learner further to respond to the feedback” (Nassaji, 2007, p. 526). For instance, in Doughty and Varela’s study (1998), phonological stress was employed to indicate the source of the problem.

Researchers have proposed that instead of adopting the dichotomic distinction of implicit and explicit feedback, the explicitness of negative feedback should be considered as constituting a continuum, as shown in Figure 3.1 (based on Loewen & Nabei, 2007, p. 362).

Loewen and Nabei proposed a feedback classification scheme that classifies feedback on two criteria: (a) provision of L2 models and pushed output, and (b) explicitness of feedback. Based on the two criteria, they divided six commonly employed feedback types in the classroom (Lyster & Ranta, 1997) into two subgroups: prompts (e.g., Ammar & Spada, 2006; Lyster, 2004; Lyster & Mori, 2006) and provide (Loewen & Nabei, 2007). The prompt group includes meta-linguistic feedback, elicitation, repetition, and clarification requests. These four types of feedback share one characteristic in common: they all prompt learners to produce output without providing L2 models. In contrast, the provide group consists of two types of feedback: explicit correction and recasts. As the name provide suggests, both types of feedback provide
models of the target form, therefore learners will not be pushed to produce output as they do with prompt.

![Diagram of feedback types](Image)

Figure 3-1. The explicitness of the feedback (Loewen & Nabei, 2007, p.362. Figure 15.1)

The subtypes of feedback under prompts and provide are ordered according to explicitness, as shown with horizontal double sided arrows in Figure 3-1. Here, explicitness is defined in terms of the identification of the error that triggered feedback. Meta-linguistic feedback is considered the most explicit form of prompt-type feedback because it explicitly identifies the source of the problem that led the NS to provide feedback. In contrast, clarification requests are on the implicit end of the continuum because learners have to infer that “the form of their utterance is responsible for the interlocutor’s comprehension problems” (Carroll & Swain, 1993, p.361). Under the term
provide, recast is closer to the implicit extreme because it does not explicitly suggest that an error was made. In contrast, explicit correction is on the explicit end of the continuum because it clearly indicates that what the student has said is incorrect (Lyster & Ranta, 1997).

3.2 Measuring Effectiveness of Corrective Feedback

The effectiveness of corrective feedback on L2 acquisition has been assessed using a range of measures in the SLA literature. These measures can be roughly divided into four types: 1. uptake and learner repair (e.g., R. Ellis et al., 2001; Lyster & Ranta, 1997; Mackey & Philp, 1998; Panova & Lyster, 2002); 2. private speech (Ohta, 2000); 3. learners’ noticing of feedback (Gass & Lewis, 2007; Mackey et al., 2000; Philp, 2003); 4. test performance (e.g., Carroll & Swain, 1993; Leeman, 2003; Sheen, 2007).

3.2.1 Uptake and Learner Repair

Uptake refers to learners’ response immediately following a NS’s corrective feedback (Loewen, 2004; Lyster & Ranta, 1998). This can be further subcategorized into two types according to the results of the correction: repair or needs repair (Lyster and Ranta, 1997). Repair refers to “the correct reformulation of an error as uttered in a single student turn” (p.49). Needs repair indicates that the utterance still contains an error, and includes six subcategories: acknowledgement (e.g., “yes” or “no” on the part of the student in response to the teacher’s meta-linguistic feedback), same error, different error, hesitation, partial repair, and off target. R. Ellis et al. (2001) employed two different terms to name the subgroups with similar definitions: successful uptake and unsuccessful uptake. Thus Uptake relates to the construct of modified output. It has been employed to examine the effects of feedback in general, and recasts in particular.
Under this approach, feedback followed by a high rate of learner uptake is considered effective, whereas that with a low rate of learner uptake is viewed as being less effective or ineffective.

Lyster and Ranta (1997), in their influential study, investigated the relationship between corrective feedback and learner uptake in four French immersion classrooms. They found that although recast is the most frequently employed feedback technique in the classroom, it leads to the least uptake compared with other types of feedback, such as meta-linguistic feedback, elicitation, and repetition. Therefore they concluded that recast is less effective than the other types of feedback in more explicit forms. In an ESL classroom-based descriptive study, Loewen (2004) found a higher rate of learner uptake compared with the study done by Lyster and Ranta (1997). He attributed the different results to the different learning contexts employed in the two studies. He argued that in ESL classrooms, L2 learners may be more ready to produce uptake as they tend to focus on linguistic forms even in meaning-focused activities, which contrasts with in immersion or content-based classes, where students tend to attend to content and meaning. Therefore, he argued that characteristics of the learning context, such as age and the previous L2 language exposure of learners, as well as classroom pedagogical focus all affect learners’ uptake. Sheen (2006) found that learner uptake following recasts also significantly relates to various characteristics of recasts, including the length of recasts, the linguistic focus (e.g., pronunciation vs. grammar), the type of change (e.g., substitution vs. addition), the mode of utterance (e.g., declarative vs. interrogative), and the number of changes (e.g., one vs. multiple).
Moreover, *uptake* is an optional discourse move; it is up to a learner to produce *uptake* or not to produce any *uptake* at all (R. Ellis et al., 2001). Therefore when learners do not produce any *uptake*, it does not necessarily mean that the feedback does not have any impact on them. Sometimes learners are just not given enough opportunities for uptake after recasts in ESL classrooms (Braidi, 2002; Oliver, 1995; 2000).

Some researchers questioned whether uptake is a reliable measure of noticing, particularly for recasts. For instance, Long (2007) argued that recasts and feedback of a more explicit nature were designed with different instructional foci: recasts can be either interpreted as a corrective feedback, or a continuous topic, therefore they are more ambiguous. In contrast, explicit feedback (e.g., meta-linguistic feedback, elicitation, repetition) aims at pushing learners to produce more output. Thus explicit feedback tends to elicit more uptake than recasts, which does not necessarily support the argument that learners notice fewer recasts than explicit feedback.

### 3.2.2 Private Speech

Ohta’s (2000) longitudinal study examined the effectiveness of recasts through *private speech*, which was recorded with small tape recorders and individual microphones clipped to seven learners of Japanese over an academic year. In her study, *private speech* was defined as an “utterance with reduced volume, lack of adaptation to an interlocutor, and lack of response from an interlocutor” (p.55). *Private speech* is viewed as “part of the dynamic internalization process, in which language development and cognitive development are mutually constitutive and proceed. For L2 learners, *private speech* serves as a cognitive tool for the ‘internalization of the L2, and
as a productive, hypothesis-testing device whereby learners think aloud as they work to solve L2 problems” (p.53).

Ohta’s study is important for the following reasons. First, in contrast to studies conducted within an interactionist framework, this study employed a qualitative discourse-analytic approach in its data analysis, which enabled the researcher to examine the efficacy of recasts from an *emic* rather than *etic* perspective. The researcher used Conversation Analysis with specific attention to overlap, pauses, as well as the levels of volume of learners’ utterances. The collected data were viewed in an *holistic* fashion. Second, besides the learner who was addressed by the teacher and received the recasts, the study also looked at the effects of recasts on learners who were not being addressed, but participated in the class as auditors. In fact, they were the ones who produced *private speech* more often than the learners who received the recasts did. The *private speech* produced by the auditors also provided counter-evidence to the notion that recasts are ineffective, since they elicited fewer responses from the learners being addressed. In this study, the auditors were found to actively participate in the class through *private speech*. As argued by Ohta, “private speech provides powerful evidence of the mental activity triggered by the noticing of contrasts between ill-formed and correct utterances. The efficacy of recasts should not be doubted based on the presence or absence of an overt oral response” (p.66).

A problem of employing *private speech* to examine the effectiveness of recasts is that learners may not be equally *talkative*. In fact, among the seven learners who participated in the study, only four of them produced *private speech* with moderate to high frequency. Ohta’s study suffered less from this drawback as it was a longitudinal
study, in which each learner could be observed over long spans of time. On the other hand, if private speech was used in a cross-sectional study, a theoretical triangulation (using multiple perspectives to analyze the same set of data), or a stimulated recall could be helpful in capturing the impact of recasts on learners who do not produce private speech frequently, but participate in the class equally actively in their mental processing (Mackey & Gass, 2005).

3.2.3 Learner’ Noticing of Feedback

Many researchers argue that corrective feedback is facilitative in L2 acquisition because it connects several mechanisms in the process, including input, internal learner capacities, output, and selective attention of learners (Long, 1996). Among these mechanisms, attention is a limited capacity system, which is required for all learning (Schmidt, 1995, Tomlin & Villa, 1994). As Gass pointed out, “One way in which the input becomes more manageable is by the learner focusing attention on a limited and hence controlled amount of data at a given point in time,” as “learners can create a set of data that allows them to move from input to output” (1997, p.8). Tomlin and Villa (1994) distinguished attention into three functions: alertness, orientation, and detection. They claimed that detection “demands more attentional resources and enables further processing of a stimulus at higher levels, such as storage and rehearsal in short-term memory” (p.296), which does not necessarily involve awareness. In contrast, Schmidt (1990, 1993, 1995) argued that awareness at the level of noticing is crucial for learning. In his well-known Noticing Hypothesis (Schmidt, 2001), Schmidt claimed that learners must consciously notice input in order for it to become intake. Hence noticing is considered requisite for learning to happen (Schmidt, 1995). Noticing here refers to a consequence of encoding “in short-term memory”, and “the detection and registration of
stimuli in short-term memory” (Mackey et al., 2000, 474). Awareness, in contrast, was assessed as the ability “to verbalize the rules of the sequence in answer to a question about a pattern underlying the stimulus presentations” (Robinson, 1995, p.300).

In the SLA literature, various retrospective methodologies have been employed in collecting noticing data, including diary studies (Schmidt, 2001), questionnaires), and uptake sheets (Warden et al., 1995). However, these methods tend to take long spans of time. As cognitive processing is claimed to occur in a short period of time, more time-efficient verbal reports have been employed, including think-aloud protocols (e.g., Leow, 2000; Rosa & O’Neill, 1999), immediate recall (Egi, 2004), and stimulated recall (Adams, 1991; Mackey et al., 2000; Slimani-Rolls, 2005).

The verbal reports can roughly be categorized into two types based on their currency of verbalization: concurrent or retrospective (Adams, 2003; Egi, 2004). In previous studies, two types of verbal reporting have been used by many researchers in collecting noticing data related to recasts: immediate recall (e.g., Egi, 2004; Philp, 2003), and stimulated recall (Adams, 1991; Egi, 2004; 2007; Nabei & Swain, 2002).

Immediate recall is a technique that can be employed to “elicit data immediately after the completion of the event to be recalled” (Mackey & Gass, 2005, p.85). This technique has been commonly used within cognitive psychology to access detection and rehearsal in short-term auditory memory. Philp (2003) employed this technique in her study to investigate various constraints preventing learners from noticing the gap between their ill-formed utterances and the target forms, after receiving recasts. In her study, learners were asked to repeat the last part of the recast they heard during a conversational turn immediately after a recall prompt signal, which consisted of two
knocks on the table. The recall prompt signal made “learners more alert to the details of that speech than they may have been ordinarily” (p.109). Egi also used immediate recall in her study, but with a different operationalization (2004). Instead of asking learners to repeat previous content from the recasts, she asked them to verbalize their thoughts about the language episodes after the recasts. Obviously, higher levels of noticing are required for learners when being asked to verbalize their thoughts, compared to simply repeating the limited items that appear in immediately preceding recasts. Although the immediate recall operationalized in Egi’s study might better capture “qualitative aspects of noticing” (2004, p.248), because of the longer time learners spent on commenting about the recasts, it might place a heavier burden on their working memories. In addition, as the knocking recall signals occurred frequently during the interaction, the flow of normal conversation was consistently interrupted. In fact, the results showed that the average length of immediate recall protocols was significantly shorter than stimulated recall. As Egi noted, it may suggest that learners wanted to keep the interruption to the communicative tasks as small as possible by trying to keep their comments concise.

In stimulated recall, participants are prompted to recall thoughts they had while performing a task or participating in an event. It is assumed that some tangible (perhaps visual or aural) reminder of an event will stimulate recall of the mental processes in operation during the event itself. Through the use of stimulated recall, “a subject may be enabled to relive an original situation with great vividness and accuracy if he is presented with a large number of the cues or stimuli which occurred during the original situation” (Bloom, cited in Mackey & Gass, 2005, p.17).
In two descriptive studies (Gass & Lewis, 2007; Mackey et al., 2000), *stimulated recall* was employed to examine the perceptions of L2 learners about feedback given on various linguistic domains, including phonology, morphology, lexis, and semantics. In the two studies, a communicative task between a learner and a NS interviewer was video-taped, and played by a second interviewer to the same learner immediately upon the completion of the activity. When watching the video, the learner was asked to pause it anytime when she/ he wished to recall any thoughts she/ he had at any particular moment in the interaction. After each interactional feedback episode, the researcher also paused the video and asked the learner to recall his/ her thoughts at the time. The goals of the studies were to examine whether learners could correctly perceive feedback with various linguistic foci, and the differences between learners with different language backgrounds (e.g., heritage vs. non-heritage learners) in terms of perceptions about feedback.

There are several benefits in employing *stimulated recall* immediately after the activity to examine the perceptions of learners concerning recasts and other feedback techniques: first, it increases the likelihood that learners trace their thoughts from short-term memory. As the memories are still fresh, they should be more close to the real thoughts that learners had at the moment of interaction. If the *stimulated recall* was conducted with a longer delay following the activity, learners may tend to say what they think the researchers want to hear because the event is not sharply focused in their memories anymore. Second, the use of the video recording serves as a strong stimulus to activate learners’ memory structures (Mackey & Gass, 2005).
On the other hand, concerns have been raised questioning the validity and reliability of the method (for a review, see Gass & Mackey, 2000). Validity includes veridicality and reactivity. Veridicality concerns the “accuracy of verbal protocols as a reflection of subjects’ cognitive processes; protocols are non-veridical if they fail to capture participants’ thoughts or include cognitive process that actually did not take place” (Egi, 2004, p.245). Stimulated recall is considered valid under the assumption that the cognitive processing of learners is directly accessible and available through their verbal reporting. However, do the verbal comments made by participants truly reflect the thought processes of participants? Researchers such as Nisbett and Wilson argued that since conscious awareness can only relate to the products of mental processes, and the processes themselves cannot be reached through introspection, learners’ verbal reporting is not based on any true introspection. Therefore it is inappropriate to use it to access learners’ cognitive processing. In contrast, the argument against Nisbett and Wilson is that the thought processes should be accessible through stimulated recall, with as “little time lag as possible between the event to be recalled and the probe” (Gass & Mackey, 2000, p.107). A second concern is reactivity, which refers to “the effect of verbalization on participants’ task performance. Verbal reports are reactive if they influence participants’ performance (Egi, 2004, p.245). The results of previous studies with regard to the reactivity of stimulated recall are conflicting. For instance, verbalizations were found either to decrease or improve learners’ performance in different studies (Leow & Morgan-Short, 2004; Rosso et al., 1989).
As pointed out by Gass and Mackey, with regard to reliability, previous studies showed that verbalization affects cognitive processes “only if the instructions require verbalization of information that would not otherwise be attended to” (2000, p.109). Therefore verbal reports that are “elicited with care and interpreted with full understanding of the circumstances under which they were obtained, are a valuable and thoroughly reliable sort of information about cognitive processes” (2000, p.109).

In addition to the validity and reliability of the instrument, several things could be done to improve the validity of these studies. First of all, external validity could be increased by recruiting a larger sample size. Second, internal validity could be increased by controlling the proficiency level of the subjects through a proficiency test, such as an Oral Proficiency Interview (OPI) before the data collection. As in Mackey et al.’s study, the number of years of prior study of English varied from as short as three years, to as long as 14 years; it is therefore hard to tell whether the results of the study would be valid for learners with various proficiency levels, or only for learners with relatively high proficiency levels given that half of the participants in the ESL group had been learning English for over ten years. In Gass and Lewis’s study, the heritage learners’ proficiency level was based solely on their course enrollment, which is also problematic as I argue in Chapter 2.

Both stimulated recall and immediate recall are important methods in eliciting noticing data on recasts and other corrective feedback. The advantages and disadvantages of the two methods were closely examined by Egi in her experimental study (2004). She argued that immediate recall has an advantage over stimulated recall in being “free from a memory decay problem” as it elicits noticing data when information
is still fresh in a learner’s STM. It also “captures the critical moment of noticing”. In terms of practicality, immediate recall requires a “short time investment for preparation and administration” (p.259). On the other hand, stimulated recall is superior to immediate recall in the following areas. First, the interlocutor in the interactional activity and the researcher in the stimulated recall session can be served by different researchers, thereby reducing the pressure on the subjects to have to avoid reporting any negative comments about the interaction to the same researcher who conducted the activity. Moreover, the researcher can clarify any uncertainties and doubts about learners’ comments on the spot. In addition, the stimulated recall is not intrusive, since it is carried on after the completion of the oral interaction. Lastly, I also find that stimulated recall requires less training prior to data collection compared with immediate recall, for both the researchers and the participants.

3.2.4 Test Performance

In classroom-based observational research, as various variables are likely to be present within the same context, it is hard to conclude whether a cause and effect correlation exists between the variable under study and the outcome. In contrast, in experimental settings, variables can be tightly controlled and manipulated. In addition, subjects can also be randomly assigned to different experimental groups. Therefore the experimental approach has an advantage over classroom-based observation in determining the correlation between the tested variable and the outcome (Mackey & Gass, 2005). Learners’ test performance has been adopted to examine the effectiveness of recasts and other feedback types in experimental studies, employing an immediate post-test (e.g., Loewen & Nabei, 2007; McDonough, 2005; Russo et al., 1989) or both immediate and delayed post-tests (e.g., Ammar, 2008; Carroll & Swain,
a pre-test was given to the participants in order to ensure group comparability prior to the treatment (Mackey & Gass, 2005). The short-term effects of treatment were measured by an immediate post-test, whereas the long-term effects were measured by a delayed post-test. A wide range of tasks have been employed for tests, including grammaticality judgment tests, sentence completion, picture prompt tests, translation tests, oral imitation tests, picture description tests, sport the differences tasks, jigsaw tasks, consensus tasks, ordering tasks, and consciousness-raising tasks (for a review see Mackey & Gass, 2006). The target linguistic structures also vary. The most frequently examined ones have been syntactical, morphosyntactical, and morphological, such as dative verbs (Carroll & Swain, 1993; Kim & Mathes, 2001), verbal predicates and particles (Nagata, 1993), indefinite articles (Muranoi, 2000), derivations of nouns from verbs (Carroll, 2001), noun-adjective agreement (Leeman, 2003), pronouns (Sanz, 2003), gender (Lyster, 2004), verb past tense (R. Ellis et al., 2006; R. Ellis, 2007), and question forms (Loewen & Nabei, 2007). The target structures in these studies have different degrees of complexities in terms of learners’ processing and operational burdens. Some structures might only require simple explicit knowledge (Leeman, 2003), while others, such as question forms (Loewen & Nabei, 2007), may require learners to acquire more complex rules.

The designing of tests is highly related to validity and reliability of the final results. The tests employed in previous studies often more likely involved grammaticality judgment tests, sentence completions, picture prompt tests, and translation tests, etc. All of these tests favor explicit knowledge rather than implicit knowledge. Therefore
studies which employed these types of tests had a higher likelihood of finding explicit feedback to be more effective than implicit feedback (R. Ellis, 2007). In order to examine the effect of feedback, employing tests that tap the implicit knowledge of learners is thus also critical. For instance, an elicited imitation test can be used to measure learners’ implicit knowledge. The assumption is that if “a given sentence is part of one’s grammar, it will be relatively easy to repeat…the sentences are typically designed to manipulate certain grammatical structures, and a person’s ability to repeat the sentences accurately is a reflection of his or her internal grammatical system” (Mackey & Gass, 2005). Caution should been used in designing appropriate instruments to test what the research intends to test (Mackey & Gass, 2005). For instance, Han (2000) suggested that grammaticality judgment tests suffer a reliability and validity problem when testing learners’ knowledge of grammatical rules. The results found in his study indicated that learners used different types of knowledge under different conditions. For instance, they tend to use their implicit knowledge under time pressure, but explicit knowledge when adequate time is given. For linguistic forms with highly complex and difficult structures, they tend to cope with intuitions as well as various strategies. However, in this small scale study only one type of grammaticality judgment test was employed. Han urged researchers to be careful before making any conclusion based on grammaticality judgment tests. Thus, employing multiple tests that examine different types of knowledge is crucial in test design. Section 3.3 will introduce several studies that employed multiple tests.
3.3 Previous Empirical Research Comparing the Effects of Implicit and Explicit Feedback

Numerous studies have been done in the past comparing the effects of different types of corrective feedback on L2 acquisition. Generally, the results of these studies showed that both implicit and explicit feedback facilitate L2 acquisition. However, in terms of which type of feedback has more advantages in L2 development, the results were mixed. Since these studies were conducted in a wide range of settings (e.g., laboratory, classroom, or computer-based interactions), and using a variety of measures (e.g., grammaticality judgment tests, sentence completion, translation tests, etc), and treatment tasks (e.g., communicative tasks, and mechanical exercises), it is not wise to rush to a conclusion about which type of feedback has a greater impact on L2 acquisition, without carefully considering the substantial differences among these studies. For instance, researchers argued that the settings of the interaction affect the effects of feedback, particularly where recasts are considered. In the classroom context, recasts are provided in an interactional context (Lyster, 1998b), therefore it may be more challenging for learners to perceive the corrective intentions of recasts as they would in a laboratory setting. Lyster (1998b; 2007) found that recasts might not function as effectively as other types of feedback in classroom settings that are more meaning-oriented than form-oriented. In laboratory settings, variables can be easily controlled, allowing for feedback to be delivered in an intensive manner focusing on the target linguistic forms.

This section will focus on 18 previous studies that compared the effects of implicit and explicit feedback conducted in three different settings: laboratory (Carroll & Swain, 1993; Kang, 2009; Kim & Mathes, 2001; Leeman, 2003, McDonough, 2005,
summarized in Table 3-2), classroom (Ammar & Spada, 2006; DeKeyser, 1993; R. Ellis et al., 2006; Havranek & Cesnik, 2001; Loewen & Nabei, 2007; Lyster, 2004; Lyster & Izquierdo, 2009; Muranoi, 2000; Sheen, 2007, summarized in Table 3-3), and computer-based interaction (Nagata, 1993; Rosa & Leow, 2004; Sanz et al., 2009, summarized in Table 3-4). The review will particularly address the test design and results in these studies.

The five laboratory studies all incorporated recasts as the form for implicit feedback. Leeman (2003) contrasted these with explicit negative evidence. McDonough (2007) was the only researcher who also used requests for clarification as an implicit feedback type. The explicit feedback also took several different forms in the studies, including meta-linguistic feedback (Carroll & Swain, 1993; 2007; Kang, 2009), explicit rejection (Carroll & Swain, 1993) and enhanced silence (Leeman, 2003). Treatments in these studies generally involved communicative tasks, which were conducted one-on-one between the interlocutor and the learner. All the studies employed pre- and post-test designs, three studies also used delayed post-tests (Carroll & Swain, 1993; Leeman, 2003; Kang, 2009). The tasks of the tests varied, but only one study employed multiple tasks during the test session (Kang, 2009), the other three only used one task. The results of these studies were mixed.

Carroll and Swain (1993) compared the effects of four types of feedback: explicit correction, direct meta-linguistic feedback, indirect meta-linguistic feedback, and recasts. The first two are more explicit, whereas the last two are more implicit. A recall task was employed in measuring the effects of the feedback, in which learners were asked to recall the linguistic items they saw during the treatment sessions. There were
two recall sessions: the first recall session was conducted immediately after the feedback session, whereas the second one happened one week later. All the experimental groups did significantly better than the control group on both recall sessions, which indicated a learning effect of all the feedback types tested in this study. However, only the direct meta-linguistic group, in which participants were told that an error was made and were provided with an explicit semantic or phonological explanation, significantly outperformed the other experimental groups. There were no significant differences among the other three experimental groups. This study was replicated by Kim and Mathes (2001). However, they only compared two types of negative feedback: the explicit negative feedback group with explanation, and the modeling plus implicit negative feedback group. The results showed no significant differences between the two feedback groups.

Leeman (2003) did not intend to compare the effects of implicit and explicit feedback, rather she was interested in the role played by negative evidence in recasts. Therefore she manipulated positive and negative evidence in the feedback employed in her study, and employed four types of feedback: recasts, negative evidence, enhanced salience of positive evidence, and unenhanced positive evidence. Among the four types of feedback, recasts were relatively more implicit, whereas the other three were more explicit. The test used in this study was a picture description task, in which learners were given two digitally altered photographs and were asked to work individually to find differences between the two photographs. The findings showed no significant differences between the four experimental groups. In addition, only the recast and enhanced-salience groups significantly outperformed the control group. McDonough
(2007) used a narrative task to measure the effects of two types of implicit feedback given on the emergence of English simple past activity verbs: recasts and requests for clarification. The results showed no significant differences between the two feedback groups, although both experimental groups did significantly better than the control group. However, the post hoc analysis showed that the clarification requests group used significantly more progressive activity verbs than the recast group. The findings suggested that clarification request may impact “several forms across developmental stages simultaneously, whereas recasts may have a more concentrated impact on a single developmental feature” (p.337).

Kang (2009) employed two tests in assessing the effects of implicit and explicit feedback on 34 English-speaking learners of Korean as a less-commonly-taught foreign language in the U.S: an untimed grammaticality judgment test and a picture description test. Implicit feedback took the form of repetitions and recasts, whereas the explicit feedback took the form of meta-linguistic comments, plus an alternative form or solution to the error. The results showed no significant differences between the two experimental groups, although both of them significantly outperformed the control group.

In classroom settings, recasts have been frequently used in the comparison of the effectiveness of prompt type (which includes clarification requests, repetition of learner error, meta-linguistic cues, and elicitation). Among these feedback types, clarification requests are more implicit, whereas the other three types of prompts feedback are more explicit. The following nine classroom studies were conducted across a spectrum of classroom settings, including immersion classrooms (Lyster, 2004), high schools (DeKeyser, 1993), postsecondary EFL classrooms (Havranek & Cesnik, 2001; Kim &
Mathes, 2001; Muranoi, 2000; Sheen, 2007), intensive elementary EFL classrooms (Ammar & Spada, 2006), and postsecondary French L2 classrooms (Lyster & Izquierdo, 2009). Among them, six studies found that implicit feedback is less effective than more explicit feedback types (Ammar & Spada, 2006; R. Ellis et al., 2006; Havranek & Cesnik, 2001; Lyster, 2004; Muranoi, 2000; Sheen, 2007). The other four found no differences between the different feedback groups (DeKeyser, 1993; Loewen & Nabei, 2007; Kim & Mathes, 2001; Lyster & Izquierdo, 2009).

Havranek & Cesnik (2001) collected, in total, 1,700 instances of corrective feedback data from L2 learners with a wide range of age and proficiency levels, out of which 12 types were identified as the most frequent combinations. The four types of correction analyzed in this study were: recasts with repetition, recasts without repetition, successfully elicited self-correction, and explicit rejection before recasts without learner repetition. It is the only study among the 18 studies reviewed in this section that examined the effects of feedback not only on learners who received it, but also on their peers who observed the feedback happening. The study employed a variety of task types, including written and spoken completion tasks, translation, correction, and read-aloud tasks. Elicited self-correction turned out to be the most efficient combinations for all learners, whereas recast without repetition was the least effective combination.

R. Ellis et al. (2006) compared the effects of two types of feedback—recasts versus meta-linguistic information—among three classes of lower intermediate EFL learners. Three different tasks were employed, including an oral elicited imitation test, an untimed grammaticality judgment test, and a meta-linguistic knowledge test. For the first two tests, grammatical and ungrammatical items were examined separately in order to
measure different types of knowledge: ungrammatical sentences provided a stronger measure of explicit knowledge, whereas grammatical sentences provided a measure of implicit knowledge. The meta-linguistic feedback group did significantly better than the recast group on the oral imitation and grammaticality judgment tasks during delayed post-tests, which were conducted two weeks after the treatments. Sheen (2007) also compared the effects of recasts and meta-linguistic feedback. The tests employed in this study included a speeded dictation test, a writing test, and an error correction test. All these tests favored explicit knowledge over implicit knowledge. The results showed that the meta-linguistic group significantly outperformed the recast and control groups on both the immediate post-test and the delayed post-test. However, the recast group did not perform significantly better than the control group.

Both of the following studies compared the effects between prompts and recasts on 118 participants who received five weeks of form focused instruction (FFI) in a French immersion program. Prompts in Lyster’s (2004) study included four types of feedback: elicitation, repetition, meta-linguistic feedback, and clarification requests. In this study, all of these four types of feedback were put under the prompts group, in comparison with a recast group. The study employed two oral tests and two written tests. The oral test included an object identification test and a picture description test, whereas the written test consisted of binary-choice and text completion tasks. The results showed that the prompts group significantly outperformed the control group on all measures. The recast group significantly outperformed the control group in some of the measures. Prompts in Ammar and Spada’s (2006) study was operationalized in three forms: elicitation, repetition, and meta-linguistic feedback. In this study, the three types of
feedback were put under the prompts group, in comparison with the recast group. The study was conducted in ESL classrooms over a four-week period. Two tasks types were employed in the study, both of which required simultaneous focus on form and meaning: one was a written meaning-oriented contextualized grammaticality judgment task, in which learners were asked to identify all errors in a story and provide the correct alternatives. The other one was an oral picture description task, in which learners had to describe a set of pictures. Both experimental groups significantly outperformed the control group on all the measures. The prompts group significantly outperformed the recast group on all the tests. In addition, Ammar and Spada also found that prompts were especially effective for learners who had lower scores on the pre-test, whereas learners with higher scores on the pre-test benefited similarly from both recasts and prompts.

Muranoi (2000) examined the impact of interaction enhancement (IE). This refers to a communicative instructional technique, in which the interaction is enhanced by means of feedback (e.g., requests for repetition and recasts), followed by two types of debriefings, during which the instructor reviews student performance in class in the learners’ first language. The learners were divided into two groups: (A) IE + formal debriefing group, with explicit grammar explanation; (B) IE + meaning-focused debriefing, in which the teacher made comments on the students’ performance in terms of accuracy in communicating the message, not in terms of accuracy of the target forms. The study employed four tasks: an oral story description task, an oral picture description task, a written picture description task, and a grammaticality judgment test.
On the delayed post-test, Muranoi found that group A who received a formal debriefing, significantly outperformed Group B, who received a meaning-focused debriefing.

Loewen and Nabei (2007) employed two tests in measuring the effects of three types of feedback: recasts, request for clarification, and meta-linguistic feedback. The results showed a learning effect of all the feedback types, but only on a timed grammaticality judgment test, not on untimed and oral production tests. And no significant differences were found among the three feedback groups.

DeKeyser (1993) studied 35 Dutch high school seniors who were learning L2 French in two classes taught by two different teachers. One teacher was asked to correct mistakes made by learners as frequently and explicitly as possible, whereas the other was asked to avoid error correction as much as possible. The treatments lasted a full school year with no particular forms targeted. DeKeyser conducted three oral tests and one written test twice, once at the beginning, and once at the end of the school year. No statistical differences were found between the two groups of students.

Lyster & Izquierdo (2009) investigated the effects of prompts and recasts, in a classroom + laboratory setting, on the acquisition of grammatical gender by 25 intermediate-level French L2 learners. Prompts in this study were operationalized as clarification requests followed by a repetition. The classroom treatment was held during a two-week period for all participants by the same instructor in a form-focused instructional unit, in which a workbook was designed to target a set of French gender endings. It was followed by a laboratory feedback treatment, in which participants completed three different oral tasks. Either recasts or prompts were given to the participants on their gender-related errors by the researcher, depending on the
feedback condition. This study employed three tests: an object-identification task, a picture-description task, an oral imitation test, a computerized binary-choice and reaction-time test. The results showed no significant difference between recast and prompt groups for any of the measures.

The following four empirical studies attempted to address the effectiveness of computer-assisted-language-learning (CALL), particularly where the question of providing feedback is concerned. Researchers argued that feedback provided through CALL is more effective since it can be provided immediately, thus it can draw learners' attention to problems in their utterance (Nagata, 1993).

Nagata (1993) compared two versions of Japanese computer-assisted-language instruction (CALI) exercises: traditional CALI (T-CALI), and intelligent CALI (I-CALI). T-CALI only provided learners feedback about what was wrong, such as “GA is not expected to be used here”. On the other hand, I-CALI not only pointed out what was wrong, but also provided detailed meta-linguistic explanations on the reason for the errors. Nagata found significant main differences between the two experimental groups in their achievement and retention for the particle errors, but not for verbal errors. Rosa and Leow (2004) examined the role of awareness in L2 development. In their study, explicit feedback was operationalized as a meta-linguistic explanation on the linguistic form selected by learners. When the selection was correct, in addition to the provision of the reason why the choice was correct, learners also received an explanation of how the targeted structure works in target language. On the other hand, when the learner made an incorrect choice, the reason for the error was provided, followed by an explanation of how the targeted structure should work. Then the learner
was instructed to try again, and make a correct selection before moving to the next item. For the implicit condition, learners only received feedback that indicated whether the answer was right or wrong, which was similar to the implicit feedback employed in Nagata's study. Two tests were employed in this study: multiple-choice recognition tests and controlled-production tests. In the multiple-choice recognition test, learners needed to choose the correct answer from four options. In the controlled-production test, learners were asked to complete sentences by providing an appropriate verb form. The results showed that learners who received explicit feedback reported higher levels of awareness of the target structures than those who received implicit feedback. In addition, learners who reported being aware of the target structure did significantly better than those who did not report being aware of it.

Sanz (2003) employed computer delivered input processing instruction, in which learners received a response from the computer of “ok” if they gave a correct answer, and “sorry, try again” if their response was incorrect. Two interpretation tests were employed in this study: a sentence completion test and a written video retelling test. Although both groups did significantly better on the post-test; no significant differences were found between the two groups.

Sauro (2009) compared recasts and meta-linguistic feedback on the English zero article among 23 high intermediate/advanced EFL adult learners. The treatment consisted of two computer-mediated collaborative writing activities, in which the learner was paired with a native English-speaking chat partner in individual chat rooms using the Virtual Classroom chat tool from Blackboard. Feedback was provided by the chat partner whenever the learner made an error with the target form. The study employed a
computer-delivered acceptability judgment test to measure the effects of the feedback. The results showed no significant differences between two feedback types for both immediate and delayed post-tests.

In sum, the 18 studies detailed revealed mixed results in terms of the effects of implicit and explicit feedback on learners’ L2 development. Some studies showed that explicit feedback was more effective than implicit feedback, especially when the study was conducted in a classroom. On the other hand, results from most of the laboratory and CALL settings found no significant differences between implicit feedback and explicit feedback. Regardless of the setting in which the research was conducted, the treatment in most of these studies only lasted for a very short period of time. DeKeyser’s (1993) study, which lasted for a full school year, was the only longitudinal study of the 18 reviewed here. No statistical differences were found between implicit and explicit feedback groups. This study reminded researchers that the time span of feedback may also affect the degree to which feedback is effective. Besides the settings of the studies, the effects of feedback were also correlated with other factors. For instance, the effects of feedback were affected by linguistic forms. Nagata found that relatively more implicit feedback significantly outperformed more explicit feedback on particles, but not on verbal predicates, which may suggest that a linguistic form may favor one type of feedback over the other. Learners’ language proficiency was also found to be associated with the effectiveness of certain types of feedback in Ammar and Spada’s study. Prompts worked more effectively for learners with lower proficiency. For learners with higher proficiency, however, both recasts and prompts seemed to work equally efficiently. This result provides some insights to the current study: since learners
with HL background usually tend to have higher proficiency compared with their non-HL counterparts, would non-HL learners also favor prompts over recasts?

In terms of test design, some of the studies only employed tests that can be considered valid for explicit knowledge. For instance, in Rosa & Leow’s study, the results showed an advantage of explicit feedback over implicit feedback. However, since the tests employed in this study only tested the explicit knowledge of learners, the validity of the results was questionable. In contrast, some of the studies employed multiple tests that can assess both implicit knowledge and explicit knowledge. The most representative example is Muranoi, who used two oral tests which measured implicit knowledge, as well as two written tests which tested the explicit knowledge of learners. Unfortunately, he did not take advantage of these tests to really compare the effects of implicit and explicit feedback.
<table>
<thead>
<tr>
<th>Study</th>
<th>participants</th>
<th>Target structure</th>
<th>Design</th>
<th>Tests</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carroll &amp; Swain (1993)</td>
<td>100 Spanish adult ESL learners (low intermediate)</td>
<td>Dative verbs</td>
<td>Groups: (A) direct meta-linguistic feedback, (B) explicit rejection, (C) recasts, (D) indirect meta-linguistic feedback, (E) control. Treatment: two feedback sessions, each followed by recall</td>
<td>Two recalls served as post- and delayed post-test.</td>
<td>All of the treatment groups significantly outperformed the control group on both recall tasks. Group A significantly outperformed other experimental groups.</td>
</tr>
<tr>
<td>Leeman (2003)</td>
<td>74 university learners of L2 Spanish (first year)</td>
<td>Spanish noun-adjective agreement</td>
<td>Groups: (A) recast, (B) negative evidence (source or problem indicated but not corrected), (C) enhanced salience with no feedback, (D) control. Treatment: Two communicative tasks: 1. Object-placement task, 2. catalog-shopping task. One-on-one with researcher.</td>
<td>Design: Pre-/Post-/delayed post-test. Task: picture description task.</td>
<td>Only groups A and C outperformed the control group on all post treatment measures. No difference between A and C.</td>
</tr>
<tr>
<td>McDonough (2007)</td>
<td>74 EFL university students (first year)</td>
<td>Simple past activity verbs</td>
<td>Groups: (A) recasts, (B) clarification requests (e.g., huh? Pardon? What? Again?), (C) control. Treatment: 1. two-way information exchange, 2. One-way information gap task, one-to-one with the researcher.</td>
<td>Design: Pre-post-test Task: a dream narration task</td>
<td>Both groups A and B significantly outperformed group C, but no significant differences shown between these two groups.</td>
</tr>
<tr>
<td>Study</td>
<td>participants</td>
<td>Target structure</td>
<td>Design</td>
<td>Tests</td>
<td>Results</td>
</tr>
<tr>
<td>---------------</td>
<td>-------------------------------------------------------------------------------</td>
<td>------------------------</td>
<td>------------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Kang (2009)</td>
<td>34 learners of Korean (with Korean-speaking parents)</td>
<td>Korean past tense</td>
<td>Groups: (A) explicit (meta-linguistic comments, alternative form ), (B) Implicit (partial or full repetition or reformulation), (C) control</td>
<td>Design: Pre-post-delayed test. Tasks: 1. grammaticality judgment, 2. picture description tests.</td>
<td>No significant difference between experimental groups on post-test.</td>
</tr>
<tr>
<td>Kim &amp; Mathes (2001)</td>
<td>20 Korean adult ESL learners (high beginners and intermediate)</td>
<td>Dative verbs</td>
<td>Groups: (A) received explicit meta-linguistic feedback, (B) received recasts. Treatment: Feedback was presented in two sessions 1 week apart each followed by production with no feedback.</td>
<td>Controlled production tasks (as in the treatment) without feedback.</td>
<td>Differences between performance on first and second production tasks were not significant. Differences between groups for increases in production were not significant. Learners expressed preference for explicit feedback.</td>
</tr>
<tr>
<td>Study</td>
<td>participants</td>
<td>Target structure</td>
<td>Design</td>
<td>Tests</td>
<td>Results</td>
</tr>
<tr>
<td>-----------------------</td>
<td>---------------------------------------------------</td>
<td>-------------------------------------------------------</td>
<td>----------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------</td>
<td>---------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>DeKeyser (1993)</td>
<td>35 Dutch high school seniors learning L2 French</td>
<td>Variety of features, predominantly morphosyntactic</td>
<td>Groups: (A) extensive explicit corrective feedback during normal class activities, (B) limited explicit corrective feedback. Treatment: 10 class periods</td>
<td>Three oral communication tasks (interview, picture description and storytelling). Fill-in-the-blank test. Tests administered twice.</td>
<td>No statistically significant differences evident between groups A and B. Learners with high previous achievement, high language aptitude, high extrinsic motivation, and low anxiety benefited the most from error correction.</td>
</tr>
<tr>
<td>Muranoi (2000)</td>
<td>114 Japanese college EFL students (first year)</td>
<td>Indefinite article to denote new information</td>
<td>Groups: (A) interaction enhancement (IE) by means of requests for repetition and recasts in communicative task + formal debriefing (explicit grammar explanation), (B) IE + meaning-focused debriefing, (C) control</td>
<td>Design: pre/immediate post/delayed post-test Tasks: 1. oral story description task, 2. oral picture description task, 3. a written picture description task, 4. grammaticality judgment test.</td>
<td>Both experimental groups outperformed the control group on both post-tests. Group A outperformed B on immediate post-test but not on delayed post-test.</td>
</tr>
<tr>
<td>Havranek &amp; Cesnik (2001)</td>
<td>207 university students specializing in English</td>
<td>Variaty of English phonological, lexical, and grammatical features</td>
<td>Data on 1700 corrective feedback episodes from normal English lessons.</td>
<td>Class-specific tests (translation, correction, reading aloud, and written and spoken completion tasks). Directed at corrected items.</td>
<td>Elicited self-correction was the most efficient combinations for all learners.</td>
</tr>
<tr>
<td>Study</td>
<td>participants</td>
<td>Target structure</td>
<td>Design</td>
<td>Tests</td>
<td>Results</td>
</tr>
<tr>
<td>------------------</td>
<td>--------------</td>
<td>------------------</td>
<td>--------</td>
<td>----------------------------------------------------------------------</td>
<td>-------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Lyster (2004)</td>
<td>148 (grade 5) 10-11-year-olds in a French immersion program</td>
<td>French grammatical gender (articles +nouns)</td>
<td>Groups: (A) received form-focused instruction (FFI) + recasts, (B) FFI + prompts (including explicit feedback), (C) FFI only, (D) control group. Treatment: workbook, providing a communicative context.</td>
<td>Design: pre/post-delayed post test Tasks: 1. object identification test, 2. picture description test, 3. written binary-choice a, 4. text completion test. Group B was only group to outperform control group on all measures (post-tests 1 and 2). Group A outperformed control group on 4 and 8 measures. Statistically significant differences were found between groups B and C not between A and B. Groups B significantly outperformed group A.</td>
<td></td>
</tr>
<tr>
<td>Ammar &amp; Spada (2006)</td>
<td>64 ESL student (Grade 6 intensive English ESL class)</td>
<td>Third person possessive determiners in English</td>
<td>Groups: (A) recasts, (B) prompts (elicitation, repetition, and meta-linguistic feedback), (C) control. Treatment: 330-495 min of communicative activities</td>
<td>Design: Pre-immediate post-delayed post-test. Task: 1. meaning-oriented contextualized grammaticality judgment task, 2. picture description task. Groups B significantly outperformed group A.</td>
<td></td>
</tr>
<tr>
<td>Sheen (2007)</td>
<td>80 EFL community college students (intermediate)</td>
<td>English articles</td>
<td>Groups: (A) recasts, (B) meta-linguistic correction, (C) control. Treatment: retell two narratives to the class.</td>
<td>Design: aptitude test, pre/post/delayed post-test. Tasks: 1. speeded dictation test, 2. writing test, 3. error correction test. Meta-linguistic group significantly outperformed Groups A and C. Recast group did not perform significantly better than C.</td>
<td></td>
</tr>
<tr>
<td>Study</td>
<td>Participants</td>
<td>Target structure</td>
<td>Design</td>
<td>Tests</td>
<td>Results</td>
</tr>
<tr>
<td>-----------------------</td>
<td>---------------------------------------------------</td>
<td>------------------------</td>
<td>-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>---------------------------------------------------------------------------------------------</td>
<td>-------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Lyster &amp; Lzquierdo</td>
<td>Classroom + laboratory study 25 French L2</td>
<td>French feminine endings</td>
<td>Groups: (A) recast, (B) prompt (clarification requests followed by a repetition). Classroom treatment: form-focused instructional unit (3 hr/ day over 2 weeks), provided for all participants. Laboratory treatment: three oral tasks, one-to-one with an interlocutor.</td>
<td>Pre-post-delayed post test 1. object-identification task, 2. picture-description task, 3. binary-choice and reaction-time measures.</td>
<td>No significant differences found between groups A and B.</td>
</tr>
<tr>
<td>(2009)</td>
<td>French undergraduate learners</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ellis et al.</td>
<td>34 private language school ESL students (low-intermediate).</td>
<td>English past tense -ed</td>
<td>Groups: (A) recasts, (B) meta-linguistic information, (C) control. Treatment: two story narrative tasks. Participants were assigned in a triads with the researcher.</td>
<td>Design: pre-/immediate post-/delayed post-test. Tasks: 1. oral imitation, 2. untimed grammaticality judgment test, 3. meta-linguistic knowledge test.</td>
<td>Group B significantly outperformed the other two groups for both the delayed imitation and grammaticality judgment post-tests.</td>
</tr>
<tr>
<td>(2006)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Loewen &amp; Nabei</td>
<td>66 EFL students at a Japanese university (Average years of studying English: 7)</td>
<td>English question formation</td>
<td>Groups: (A) Recasts, (B) Clarification request, (C) Meta-linguistic feedback, (D) No feedback, (E) control. Treatment: 1. a spot-the-difference, and 2. a guess-the-storyline task, one-to-one with the researcher.</td>
<td>Design: pre/post-test Tasks: 1. timed grammaticality judgment test, 2. untimed grammaticality judgment test, 3. oral production task.</td>
<td>Groups A, B, and C significantly outperformed groups D and E only on timed grammaticality judgment test, but not on untimed grammaticality judgment and oral production tests.</td>
</tr>
<tr>
<td>(2007)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study</td>
<td>Participants</td>
<td>Target structure</td>
<td>Design</td>
<td>Tests</td>
<td>Results</td>
</tr>
<tr>
<td>------------------</td>
<td>---------------------------------------------------</td>
<td>-------------------------------------------------------</td>
<td>----------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------</td>
<td>------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Nagata (1993)</td>
<td>32 second-year university learner of L2 Japanese</td>
<td>Japanese passive structures; verbal predicates and particles</td>
<td>Groups: (A) feedback indicating what was missing or not expected, (B) received same feedback + meta-linguistic explanations.</td>
<td>Written test using same format as treatment task.</td>
<td>Group B significantly outperformed group (A) on particles but not verbal predicates. Learners expressed preference for meta-linguistic explanation.</td>
</tr>
<tr>
<td>Rosa &amp; Leow (2004)</td>
<td>100 adult university learners of L2 Spanish enrolled in advanced courses</td>
<td>Contrary to the fact conditional sentences in the past</td>
<td>Groups: (A) metalinguistic explanation + opportunity to try again if incorrect, (B) implicit feedback indicating whether the answer was right or wrong, (C) control group. Treatment: input-based jigsaw task characterized by task essentialness</td>
<td>Design: pre/immediate post/ delayed post-test. Tasks: 1. multiple-choice recognition tests, 2. written controlled production tests.</td>
<td>Learners who received explicit feedback reported higher levels of awareness than those who received implicit feedback. In addition, learners who reported being aware of the target structure did significantly better than those who did not report being aware of it.</td>
</tr>
<tr>
<td>Sanz (2003)</td>
<td>28 first-year university learners of Spanish</td>
<td>Position of clitic pronouns between object and verb</td>
<td>Groups: (A) explicit metalinguistic feedback, (B) implicit feedback. Treatment: Computer-delivered input processing instruction without prior explicit instruction.</td>
<td>Design: pre/post-test. Interpretation tests. Task: 1. sentence completion. 2. written video retelling.</td>
<td>Both groups significantly increased their abilities to interpret the O-clitic V sentences, but no significant differences showed between the two groups.</td>
</tr>
<tr>
<td>Sauro (2009)</td>
<td>23 high intermediate advanced adult ESL learners.</td>
<td>English zero article</td>
<td>Groups: (A) Recasts, (B) meta-linguistic feedback, (C) control. Treatment: computer-mediated collaborative writing activities, paired with a native English-speaking interlocutor.</td>
<td>Groups: (A) recasts, (B) meta-linguistic feedback, (C) control. Treatment: computer-mediated collaborative writing activities, paired with a native English-speaking interlocutor.</td>
<td>No significant advantage for either feedback type on immediate or sustained increases, the meta-linguistic group showed significant immediate increases relative to the control condition.</td>
</tr>
</tbody>
</table>
3.4 CHL Learners, Implicit and Explicit Feedback

CHL learners usually use Chinese when communicating with family members, such as grandparents or parents, who normally do not speak English (Dai & Zhang, 2008). When speaking with siblings and Chinese friends, CHL speakers tend to switch to English. Some of them even feel embarrassed to speak their HL with people other than their family members, because they consider Chinese a language only spoken at home. The shyness toward their HL suggests that their acquisition and maintenance of their HL is in a “vertical and reciprocal intimate relation” (Dai & Zhang, 2008, p.41), and is restricted to an informal setting with limited topics of daily life. Moreover, many CHL learners speak a non-prestige dialect, which in many cases is unintelligible to Mandarin. Finally, due to their different levels of exposure to the HL, the language abilities of CHL learners vary from individual to individual, which leads to tremendous variations in their linguistic knowledge and abilities when beginning Chinese classes at college or university (Hendryx, 2008). However, as a group, CHL learners show some consistent patterns: although having developed a considerable level of fluency in their oral speech skills, they tend to have a narrow range of lexical and syntactic alternatives (Valdès, 2001). In addition, the majority of CHL learners tend to have high-level receptive skills (especially listening skills) but low-level productive skills (e.g., speaking), which often resemble other HL learners (Valdès, 2001).

The imbalance between the receptive and productive skills observed in CHL learners reminds us of learners in Swain’s well-known study on Canadian French immersion programs (1985; 2005). Both learners in the two CHL learner groups and French immersion program learner groups speak a dominant language, as well as a second language which they started to learn from a young age. Learners in both groups
are able to access an abundance of comprehensible input either in the immersion classroom or at home. Like the CHL learners, the immersion learners were also found to be weak in their productive skills (speaking and writing), although they were deemed native-like in other language areas. Swain concluded that what the immersion learners lacked was being pushed to produce modified output. She argued that modified output is necessary for learners’ acquisition because:

“while attempting to produce the target language vocally or silently (sub-vocally), learners may notice that they do not know how to say (or write) precisely the meaning they wish to convey...Under some circumstances, the activity of producing the target language may prompt second language learners to recognize consciously some of their linguistic problems: It may bring their attention to something they need to discover about their second language” (Swain, 2005, p. 474).

Similar to learners in Swain’s study, CHL learners also may not have enough opportunities to be pushed to produce modified output. Although many of them grow up in a Chinese-speaking family, the conversation between parents and children is often a mixing of English and Chinese. Usually the parents speak in Chinese, but the children respond in English. For these learners, comprehending input is easier as they can understand its meaning without precise knowledge of morphology and syntax. On the other hand, in order to produce output, they need to focus their attention on the forms in their inter-language and tap their language competence. When providing corrective feedback to CHL learners, the following questions need to be considered: which type of feedback is most likely to help learners notice the corrective intention? Which type of feedback can best push CHL learners to make modifications and produce modified output?
The fundamental difference between CHL learners and non-CHL learners lies in their internalized language system: CHL learners have already developed an implicit linguistic knowledge system of Chinese even before starting their formal Chinese study at schools, which is attributed to their earlier exposure to this language from family members during their childhood (Valdés, 2005). In contrast, non-CHL learners are not equipped with such a knowledge system beforehand, and are more likely expected to construct it solely through formal classroom instruction.

In terms of explicit knowledge, however, HL learners are not superior to their counterparts. For instance, although they seem to know the language and are able to use a set of internalized grammatical rules, they do not necessarily have the metalanguage to talk about the grammatical system itself, and may also have difficulties learning grammar rules that are commonly taught in foreign language or second language classrooms (Valdés, 1995).

Non-HL learners usually initiate their target language study by obtaining explicit knowledge through explicit instruction, and later only select portions of the explicit knowledge that they feel ready to incorporate into their interlanguage system. This could possibly be converted into implicit knowledge (Han, 2003). As CHL learners have already developed an implicit knowledge system, they would need to convert the knowledge system in an opposite process to that of the non-CHL learners.

Having discussed the profile and internal knowledge system of CHL learners, I propose the following working definition of explicit feedback in my study: information or comments which a. clearly indicate that an error has been made; b. use meta-linguistic language to point out the source of the error. On the other hand, implicit
feedback refers to information or comments which a. do not overtly indicate that there was an error; b. do not interrupt the flow of communication. Explicit feedback in my study is provided in the form of a meta-linguistic technique, whereas implicit feedback takes the form of recasts as shown in examples 3-1 and 3-2 below.

Example 3-1 Explicit feedback
010 NNS: *我 买 四 笔。
   Wo mai si bi
   1sg buy four pen
   I bought four pens.

011 NS: 我们不说四笔。你应该用一个量词。
   We don’t say sibi, you should use a measure word here.

In Example 3-1, the learner produces a problematic sentence in 010, which triggers the feedback by the NS in 011. This feedback contains two parts: the first part highlights the problematic form by repeating the learner’s original word, which indicates that the learner made an error. The second part provides a with a meta-language hint with regard to the correct form that should be used here.

Meta-linguistic feedback as defined in my study may be perceptually more salient to CHL learners for the following reasons. First, it has been estimated that HL learners have the potential ability to acquire 80%-90% of grammatical rules that govern words, phrases, and sentences (Ming & Tao, 2008). However, certain problems were also found to persist in their grammatical competency despite years of learning the language (Kim, 2008). These problems will not automatically disappear solely from the exposure to positive evidence. For them “some kind of structural analysis of input has to occur” (Han, 2003, p.134). Moreover, compared with English-speaking non-HL learners, although HL learners share similar sentence processing problems in the language, they display greater comprehension ability in listening. This superiority in comprehension by
HL learners is due to their stronger semantic knowledge and contextual skills, but not to the in syntactic structure processing skills (Kim, 2008).

“Environmental support in the form of comprehensible input is necessary for language learning, but insufficient for learning certain specifiable aspects of an L2. Paradoxically, comprehensible input may actually inhibit learning on occasion, because it is often possible to understand a message without understanding all the structures and lexical items in the language encoding it, without being aware of not understanding them all” (Long, 1996, p. 425).

Thus the meta-linguistic hint provided in meta-linguistic feedback may push learners to correct their errors through searching their implicit linguistic system. Second, it draws learner’ attention from the meaning to the form. Previous studies found that HL learners tend to focus on semantic issues rather than other linguistic features. Therefore during the interaction, HL learners are more likely to attend to meaning rather than form (Gass & Lewis, 2007, Kim, 2008). On the other hand, although HL learners have developed an implicit knowledge system of their HL from the exposure at home during their childhood, they tend to lack explicit meta-linguistic and rule-based knowledge which would possibly help them notice the gaps between their interlanguage and the target forms (Kim, 2008). Therefore they need consistent corrective feedback from the teachers to draw their attention to the forms of the language, rather than meaning when engaging in interaction (Kim, 2008). Based on all of these factors, meta-linguistic feedback stands out compared to other feedback techniques since the meta-linguistic information provided in the feedback not only helps learners figure out the correct linguistic form, but also contributes to the construction of learners’ explicit knowledge system. Currently in North America, many post-secondary Chinese programs are designed for either foreign language learner, or native Chinese learners who need little explicit instruction of grammar and pronunciation (He, 2008). However, neither of these
tracks may be truly effective for CHL learners who may need less explicit instruction than foreign language learners, but more explicit instruction than the native speakers.

Example 3-2 Implicit feedback
020 NNS: *我 买 四 笔。  
Wo mai si bi  
1sg buy four pen  
I bought four pens.

021 NS: 你 买 了 四 枝 笔?  
2sg mai le si zhi bi  
You buy PFV four CL pen  
You bought four pens?

022 NNS: 四 枝 笔。  
Si zhi bi  
Four PFV pen  
Four pens.

In the current study, implicit feedback is operationalized in the form of recasts, for the following reasons. First, although general agreement has been made that recasts can add material to a child’s grammatical repertoire during L1 acquisition, disagreement remains on whether it could also provide negative evidence for children on their ungrammatical utterances (Nicholas et al., 2001). Despite this, in SLA classrooms, recast is the most frequently used corrective feedback form by L2 teachers (Oliver, 1995; Long, 2007; Lyster and Ranta, 1997; Sheen, 2004). Beside corrective recasts, teachers also frequently use noncorrective recasts to show approval toward learners’ responses. The dual-functions of recasts usually cause ambiguity and make recasts more implicit for second language or foreign language learners. To further complicate the issue, it remains unclear whether HL learners are relatively more sensitive toward recasts because of their previous language exposure. That is, they have noticed the dual-functions served by recasts. However, from time to time, they may not process adequate knowledge to accurately interpret which function a particular recast serves. All
these uncertainties may make recasts more implicit for CHL learners. Second, recasts are less likely to be perceived as corrective feedback compared with other types of corrective feedback because of their multiple discourse functions. For CHL learners, as they have more experience communicating with their family members in HL, they tend to regard language exercises in the classroom as interactional conversation (Gass & Lewis, p.99), especially when the correction is embedded in a complete rather than in a partial reformulation. In addition, their broader implicit knowledge gives them a privilege over non-CHL learners. That is, they do not have to know every detail in the utterance before they can comprehend. Therefore they tend to be less sensitive to the alternative forms provided by NS. Lastly they tend to treat a NS’s reformulation as a yes/no question or as a continuation of the conversation, since they are not pressed to produce output.

In sum, due to their unique language exposure to Chinese at home, CHL learners develop an internal knowledge system that differs from non-CHL learners. CHL learners have developed strong implicit knowledge, but much weaker explicit knowledge. Consequently, their reaction toward implicit and explicit feedback may differ from non-CHL learners who have the opposite distribution of implicit and explicit knowledge. Thus the findings from previous studies on the effects of implicit and explicit feedback, which mainly focused on FL/SL learners, may not apply to CHL learners. However, to the best of my knowledge, little empirical research has been done in this area so far. This gap in the literature motivated the research questions of the current study, which I will present in Chapter 4. Assuming that HL learners had indeed developed a certain sensibility
toward recasts, would they form a new perception toward recasts in the classroom after observing the teacher for a certain period of time?
CHAPTER 4
RESEARCH QUESTIONS AND METHODOLOGY

Drawing from previous research on interaction, implicit and explicit feedback, and noticing, as well as HL and non-HL learners’ language acquisition, the current study aimed at investigating the effects of feedback in relation to feedback type, as well as learners’ language background. This chapter will present the research questions and hypotheses as well as the details of the methodology employed to answer the research questions in this study.

4.1 Research Questions

There are three research questions in the current study.

RQ 1: Do CHL learners perceive both explicit and implicit feedback on Chinese classifiers more accurately than non-CHL learners?

H 1: Learners with different language backgrounds would have different perceptions about the feedback provided during the interaction. As HL learners develop language sensitivity and intuitions through exposure to the HL in a natural language setting at home, they can comprehend the meanings without having to know every detail in the utterances. Consequently, they tend to focus on meaning rather than on linguistic form when engaging in a conversational interaction. The findings in previous studies suggested that HL background affected learners’ perception of the feedback: learners with HL background showed higher accuracy in perceiving feedback on semantic and lexical items compared with their non-HL counterparts. In contrast, non-HL learners were more likely to perceive feedback on semantic and lexical items as no content, a term which refers to something related to the content of the conversation between them and the interlocutor (Gass & Lewis, 2007; Kim, 2008). Since the target
linguistic form (Chinese classifiers) in the current study could be both semantic and lexical (Zhang, 2007), it was hypothesized that CHL learners would perceive both explicit and implicit feedback on Chinese classifiers more accurately than non-CHL learners.

**RQ 2**: Will explicit feedback work more effectively than implicit feedback for CHL and non-CHL learners’ language development?

**H 2**: Learners’ language proficiency level may impact the effects of different types of feedback. Ammar and Spada (2006) found that prompts—which included meta-linguistic information, repetition, and elicitation—worked more effectively than recasts for learners with lower proficiency, but not for learners with higher proficiency. As CHL learners usually have higher proficiency than their non-CHL counterparts, particularly in speaking and listening skills, it was hypothesized that explicit feedback would prove more effective than implicit feedback in non-CHL learners’ L2 development, but not in CHL learners’ HL development.

**RQ 3**: Does feedback type or language background affect the increase in CHL and non-CHL learners’ performance on a) an oral imitation test, b) an untimed written cloze test?

**H 3**: The two tests employed in the current study were designed to measure different types of knowledge. On the oral imitation test, learners are more likely need to tap their implicit knowledge to perform their oral production under time pressure. On the other hand, on the written cloze test, learners have time to access their explicit knowledge since the test is untimed. In addition, their performance is determined by
their knowledge of Chinese characters as a full score for each target linguistic form is given only when it is written in Chinese characters.

HL learners have a unique internal language knowledge system that sets them apart from non-HL learners in the process and development of language acquisition (Gass, 1997). Before receiving formal language instruction at school, HL learners have usually already developed their implicit knowledge, but lack explicit knowledge as well as literacy skills. In the case of CHL learners, they have no advantage over non-CHL learners in their learning of Chinese characters. In addition, previous research also found that HL learners performed more like native speakers than the non-HL learners on their oral productions, but more similarly with non-HL learners on their written productions (Montrul et al., 2008).

Therefore the following results are expected: a) CHL learners and non-CHL learners benefit from feedback in a different way; b) CHL learners outperform non-CHL learners on the oral imitation test, but not on the written cloze test. Thus it was hypothesized the feedback type and learners’ language background affect their increase in performance on the oral imitation test, but not on the written cloze test.

4.2 Methodology

4.2.1 Operationalizations

Feedback. Following R. Ellis et al. (2006), explicit feedback was in the form of meta-linguistic information, while implicit feedback was in the form of recasts. In order to minimize the perceptual salience of the targeted linguistic items, feedback was provided on any errors made by learners, and not limited to classifier-related errors only.

Following Lyster and Ranta (1997), meta-linguistic feedback was operationalized as “comments, information, or questions related to the well-formedness of the learner's
utterance, without explicitly providing the correct form" (p.47). An example of metalinguistic feedback taken from the data collected for the current study is given in 4-1. Transcription conventions are shown in Appendix A.

Example 4-1 Meta-linguistic feedback
NNS: *我买了一衬衫。
Wo mai le yi chenshan.
1sg buy PFV one shirt.
I bought one shirt.
NS: You need a measure word for clothing ← Meta-linguistic feedback
you wear on the top.

As shown in this example, meta-linguistic feedback includes a provision of meta-linguistic information to indicate the error made by the learner by highlighting the nature and the characteristic of the TL form. In this example, the feedback pointed out which type of words needed correcting in the learners’ problematic sentence. In addition, it also indicated the specific semantic association between the target classifier and the related head noun.

Recast was operationalized as the NS interlocutor’s reformulation of all of the problematic utterances produced by learners. In order to make recasts as implicit as possible, the corrective part in recasts was not stressed or realized with higher pitch, in order to avoid any particular salience (Leeman, 2003). Moreover, the recasts of learners’ problematic utterances were always in the form of a complete sentence (Sheen, 2006).

Example 4-2 Recast
NNS: *我买了一条衬衫。
Wo mai le yi tiao chenshan.
1sg buy PFV one CL shirt
I bought one shirt.
NS: 你买了一件衬衫。
Ni mai le yi jian chenshan. ← Recast
You bought a shirt.

In Example 4-2, the same error as in Example 4-1 triggered the recast provided by the NS. When providing the recast, the NS reformulated the learner’s original sentence with a complete sentence 你买了一件衬衫 Ni mai le yi jian chenshan in either a declarative or an interrogative form^1.(see Example 4-3). No additional features (e.g., stress, intonation, and segmentation) that could signal the error were added to the recast.

**Perception of feedback.** Following Mackey et al. (2000), as well as Gass and Lewis (2007), learners’ accurate perception of the provided feedback was operationalized as any verbal comments that indicated that the learner had noticed, or paid attention to, feedback on the classifier-related error at the time.

A typical example of accurate perception of feedback is shown below:

Example 4-3

NNS: *我 买 了 一件 帽子。
   Wo mai le yi jian maozi.
   1sg buy PFV one CL hat.
   I bought one hat.

NS: 你 买 了 一顶 帽子?
    Ni mai le yi ding maozi.
    2sg buy PFV one CL hat.
    You bought a hat?

NNS: 我 买 了 一顶 帽子.
    Wo mai le yi ding maozi.
    1sg buy PFV one CL hat.
    I bought one hat.

Recall data
I didn’t really know how to say one hat. I was not sure which measure word^2 I should use for “hat”. And she (the NS) corrected me.
In example 4-3, the learner clearly verbalized that the NS corrected his error on the classifier.

4.2.2 Design

In the literature, the effects of corrective feedback have been investigated from both classroom-based observation (Panova & Lyster, 2002; Sheen, 2004) and laboratory-based research (Ayun, 2004; Han, 2002; Leeman, 2003). Each approach has its advantages: classroom-based research can “enhance our understanding of how to implement effective ways of improving learners’ second language skills,” while laboratory studies “can provide more tightly controlled environments in which to test specific theories about second language development” (Mackey & Gass, 2005, p.186). However, in classroom-based research, it is hard to determine a cause-and-effect relationship between the investigated independent variable(s) and the dependent variable(s) since all the variables are present in the same context simultaneously without any control. This disadvantage could be overcome in an experimental or quasi-experimental study, in which independent variables are tightly controlled and manipulated. Thus the results are more straightforward and reliable. Since the goal of the current study was to investigate the association between feedback types and the learning outcome by learners with different language backgrounds, an experimental approach in which different variables (e.g., feedback type, learner group) could be controlled was desirable.

The current study employed a pre-test-post-test randomized control group design (see Figure 4-1). By randomly assigning 64 participants to experimental and control groups, it was assured that there were no biases in the assignment of subjects. Forty-seven learners in experimental groups participated in two treatments, during each of
which they engaged in a conversational interaction with a NS of Mandarin. Feedback was provided by the NS on any errors made by the learners during the interaction. Learners in each experimental group received feedback in only one form, either recast or meta-linguistic, depending on the group to which they were assigned. Their short-term language development was measured by pre- and post-tests. Further details about the treatment and test materials will be discussed in section 4.2.5. Learners’ attention to feedback was measured by stimulated recall data, which was collected on the last day of the experiment. An interview was conducted to investigate the HL learners’ language experience and home language background. Non-HL learners did not participate in the interview. Seventeen learners in a control group took only the pre- and post-test without participating in the treatments. Their scores on the two tests were compared with those of the experimental groups, to ensure that any improvements made by learners in the experimental groups were due to the treatments, rather than to test effect.

The pre-test was used to ensure “comparability of the participant groups prior to their treatment” (Mackey & Gass, 2005, p.149), while the post-test was employed to measure the treatment effect. In designing the pre- and post-test, the following issues were carefully considered. First, as the current study intended to look at the effects of feedback on each classifier, the same set of 33 sentences was used in both tests in order to ensure comparability between the two tests. In order to minimize practice effects, however, the sentences were randomly arranged in each test. Second, in order to draw different types of language knowledge from the learners, both the pre- and the post-test used two different testing instruments: an oral imitation test and a written test.
The former was used to measure learners’ implicit knowledge, while the latter was employed to assess their explicit knowledge.

The experiment was conducted on three consecutive days. The short testing period was chosen due to the characteristic of the targeted linguistic forms: the association between a classifier and available head nouns is arbitrary and conventional,
but irregular, and not rule-governed. Frequent input consisting of classifiers that are presented with associated head nouns makes the acquisition of these classifiers easier. In addition, all the classifiers employed in the current study were relevant to objects or animals usually seen in daily life. Therefore HL learners have far more opportunities to have exposure to these classifiers compared with non-HL learners. In other words, the longer the testing period lasted, the more likelihood HL learners could hear these classifiers at home when conversing with their family members. In contrast, since non-HL learners only had access to Mandarin in the classroom, it was unlikely for them to have equal opportunities to have exposure to the classifiers. In order to control for possible external input, the current study employed a two-day testing period.

The treatments were both video- and audio-recorded. The video-recording of the treatments were edited into short clips, each of which contained one feedback episode. The video clips were used as stimuli during the stimulated recall session. The stimulated recall was used to elicit the noticing data on the provided feedback from the 47 learners in the experimental groups. During the stimulated recall session, learners were asked to verbally report their thoughts after watching each video clip which contained a feedback episode taken from the two treatments. They could ask the interviewer to pause the video anytime in between when they wished to share any thoughts at any particular moment during the interaction. (See section 4.2.5.4 for more details.)

4.2.3 Participants

**CHL and non-CHL learners.** A total of 93 volunteers initially participated in the experiment, however 29 of them were excluded from the study and their data will not be reported here. Among them, four non-HL learners were excluded because they had a
much longer length of formal Chinese study than the other non-HL participants. Four non-HL learners were excluded because they were heritage speakers of Spanish \( (n=4) \). Although their HL is not Chinese, their HL experience might provide them some advantages over the English monolinguals. Seventeen HL learners were excluded because their home dialect was a dialect other than Mandarin (Cantonese=12; Fujian=3; Wenzhou=1; Shanghai=1). Four HL learners withdrew from the experiment in the middle due to a scheduling conflict. Thus a total of 64 (male=35; female=29) learners were left in the final pool of participants. Among them, there were 23 HL learners and 24 non-HL learners in the experimental groups; eight HL learners, and nine non-HL learners were included in the control group.

When the data collection was conducted, 15 participants were enrolled in the Heritage Class, which is equivalent to a second-year level traditional Chinese class. Twenty-one participants were enrolled in the third year class and 28 were enrolled in the second year course. Their ages ranged from 18 to 23, with an average age of 19.9 \( (SD=1.20) \). The average length of formal Chinese study was 10.9 months for CHL learners \( (SD=6.8) \); 19.4 months for non-CHL learners \( (SD=7.9) \). The average length of stay in China was 1.3 months \( (SD=8.0) \). All participants completed an informed consent form (see Appendix B), and agreed to receive 0.3-0.5% extra points for their final class grade from their instructors upon the completion of the study.

Why were only HL learners with Mandarin L1 and non-HL learners with English L1 involved in this study? Why were learners at high-intermediate and advanced level chosen for the experiment? These two questions will be answered below.
In Fall 2008, I conducted a preliminary survey among approximately 300 learners who enrolled in ten classes in the Chinese program at UF. Out of 267 returned responses, 79 learners were identified as CHL learners. Among them, 40% of learners reported that they grew up in a Mandarin speaking family, followed by 32.5% of learners who reported Cantonese as their family language. The rest of learners reported various dialects spoken at home, such as Shanghai, Fujian, Wenzhou, etc. Based on the survey results, I initially proposed to include both Mandarin and Cantonese speakers in the experimental group. The final decision of excluding HL learners with dialects other than Mandarin from the experiment was made based on the following two reasons. First, only 12 Cantonese speaking HL learners volunteered themselves to participate in the research, which was much lower than the originally proposed number of 30.

The second reason was due to a theoretical consideration. Although the majority of Mandarin classifiers are also available in other dialects, there are variations among different dialects. Since Cantonese was the second most commonly-spoken dialect by the HL volunteers in the current study, I will use Cantonese as an example to make a comparison with Mandarin. Mandarin and Cantonese share most of the same classifiers. However, for a same head noun, Mandarin and Cantonese may apply different classifiers. For instance, for the head noun “needle”, 针 zhen, the classifier used in Mandarin is 根 gen, while 眼 ngaan is used in Cantonese. In addition, even though a classifier is used in both dialects, the semantic meaning may differ. For instance, 只 zhi is the classifier used in Mandarin to refer to small animals such as birds, mice, and cats. It also classifies one of a pair of small objects, such as shoes, gloves,
and socks. On the other hand, 只 jek in Cantonese, sorts not only the small animals listed above, but also large animals, such as horses and oxen. It also refers to round objects, such as eggs and teeth. Even pans, windows, and phonograph records also use it as the classifier, although this would be considered incorrect in Mandarin.

Besides the differences in the use of classifiers, the most important difference that sets Mandarin and Cantonese apart is the phonology, given that the two languages are not mutually intelligible phonologically in the first place. The distance between them is as much as French differs from Spanish, or Swedish from German. The two languages have a different consonant and vowel system. Some consonants and vowels can only be found in Mandarin but not in Cantonese, and vice versa. For instance, the dental stops [z], [c], and retroflex stops [zh], and [ch] in Mandarin are particularly difficult for Cantonese speakers to learn. In addition, their tonal system is also different. Mandarin has only four tones, whereas Cantonese has at least six distinctive tones. For instance, the pitch value in Mandarin rising tone is 35; however, Cantonese has 35 or 25 for high rising, and 23 or 13 for low rising (Argus & Matthews, 1991).

Thus in the current experiment, HL learners with dialects other than Mandarin were excluded in order to reduce the extra variables that may have influenced the effects of the treatments.

Non-HL learners with English L1 were chosen for the experiment due to the following concern. As all the HL learners participating in the current study received their formal education in English-dominant mainstream schools, their dominant language was English. In fact, all 23 HL learners interviewed by the researcher considered English to be their native language. Previous research showed that the L1 variable was crucial in
determining the variability of both HL and non-HL learners. HL learners showed similar processing strategies as L2 learners when their dominant language was the same as the L2 learners’ L1 (Kim, 2006). Thus it would be interesting to compare the linguistic behavior of members in these two learner groups.

Learners at high-intermediate and advanced levels were chosen due to the following concerns: first of all, as discussed in Chapter 2, HL learners usually tend to have strong speaking and listening skills even with very little formal classroom instruction. In order to ensure the communicative activities adopted in the treatment appropriate for learners in both groups—not too easy for CHL learners on the one hand, but also not too challenging for non-CHL learners on the other hand—non-CHL learners need to reach a proficiency level that allows them to carry on communication with the NS. Secondly, since the majority of the target classifiers had already been introduced in the textbook at the beginning and intermediate level, all the non-CHL learners should have already acquired, or at least encountered all or the majority of the target classifiers from their textbook through formal classroom learning. For CHL learners, although they may not have acquired all the classifiers in the classroom as some of them skipped the beginning level, they were more likely to have heard all of these classifiers that are frequently used in daily life. Third, for HL learners, after at least one year of formal instruction at the university level, their explicit knowledge of Mandarin, which started to grow at earlier years of their life through language exposure at home, was re-acquired. Thus, the difference in the performance between them and the non-HL learners was more likely due to their HL experience, which the current study was mostly eager to explore.
**CHL and non-CHL learners’ profiles.** The following three learner profiles represent two types of CHL learners and the non-CHL learners in the current study. For the two CHL learners, learner WKW represents learners who were born in the U.S., whereas learner YHX represents learners who were foreign-born, but emigrated to the United States at young age. These two profiles briefly illustrate how these two types of CHL learners developed their HL before they took formal college-level Chinese courses.

**WKW.** WKW was born in the U.S. His parents spoke only Chinese with him at home until he attended kindergarten at age five. Besides his parents, he also practiced his Chinese with his grandfather when the latter visited the United States. He learned many Chinese classical poems from his grandfather. There was a clear change before and after he started kindergarten: before attending kindergarten, his mother read Chinese stories to him. He also watched Chinese TV programs; after attending kindergarten, his dominant language rapidly shifted from Chinese to English. He still watched some Chinese TV programs, such as the TV series *The Journey to the West, The One Hundred and Eight Heroes*, but he could only understand the plots, not the details. Worrying that he might totally forget his Chinese, his parents sent him to a community Chinese language school when he was seven years old. He went there every Sunday for a three-hour class until he was sixteen. At that time he felt learning Chinese was useless since it was not commonly used in this country. Later when he took postsecondary Chinese classes, he started to realize that he was re-acquiring many words he learned when he was young. When the data was collected, he had been taking Chinese course at the University of Florida for eight months.
YHX. She was born in Taiwan, and came to the U.S. when she was seven. Before moving to the U.S., she spoke Chinese with her parents and grandparents at home. She learned approximately two hundred Chinese characters in kindergarten, as well as in elementary school, where she studied for half a year. When she had just come to the U.S., her parents asked her to copy Chinese characters from a dictionary every day, but she quickly gave up. Between age seven and age nine, she attended a Chinese community language school every Sunday. Although her English overtook her Chinese and became dominant at around age ten, Chinese has always been part of her life: she speaks Chinese with her parents at home; she enjoys going to Chinese Karaoke with her Chinese friends; she also watches Chinese TV programs with her parents, and she estimates that she can understand 80% of them. On the other hand, although there are many Chinese books and magazines at home, she almost never bothers to read them due to her limited literacy skills. When the data was collected, she had been taking Chinese course at the University of Florida for eight months.

Learner ZK represents the non-CHL learners in the current study, who have been learning their Chinese mainly through formal classroom education at a postsecondary level.

ZK. Learner ZK is a native English speaker, who was born in the United States. He had been taking Chinese course at the University of Florida for 15 months when the data was collected. He never visited China, and his entire skills in Chinese came from classroom learning. He had Chinese language partners, with whom he sometimes practiced his Chinese. But basically he did not have many opportunities to speak
Chinese outside the classroom. He also tried to watch some Chinese movies, cartoons, or TV programs, but he admitted that he could only pick up some words occasionally.

**Interlocutor.** The interlocutor of the task-based interaction was served by the researcher, who is a female native Mandarin speaker with 13 years experience teaching Chinese as a second language. She started to teach undergraduate Mandarin courses at UF in 2007. In order to avoid sampling bias, all the students whom she was teaching when the data was collected were excluded from the experiment at the subject recruitment stage. The same interlocutor elicited the stimulated recall data and conducted the interview. She was in her late 30s when the data was collected.

### 4.2.4 Target Linguistic Items

**Chinese classifiers.** Previous studies have investigated feedback effects in various linguistic domains, such as syntactical, morphosyntactical, and morphological forms, in the forms of dative verbs (Carroll & Swain, 1993; Kim & Mathes, 2001), verbal predicates and particles (Nagata, 1993), indefinite article (Muranoi, 2000), derivations of nouns from verbs (Carroll, 2001), noun-adjective agreement (Leeman, 2003), pronouns (Sanz et al., 2009), gender agreement (Lyster, 2004), verb past tense formation (N. Ellis & Larsen-Freeman, 2006; R. Ellis, 2007), and question formation (Loewen & Nabei, 2007). The target linguistic items in these studies differed in degree of complexity in terms of learners’ processing operation burden. Some structures required simple explicit knowledge, while others required more complex rules. Thus the choice of linguistic structures relates to the results of the research, and needs to be made with careful consideration.

The current study chose Chinese numeral classifiers as the linguistic item, motivated by theoretical, methodological, and pedagogical grounds. To my knowledge,
no studies on feedback effect have been done in the past that chose classifiers as the target structure.

Before discussing the theoretical grounds of the target linguistic form, I will first provide a brief introduction to Chinese classifiers. Chinese is a numeral classifier language. Chinese classifiers can be subcategorized into three types: nominal, verbal, and compound classifiers. The current study chose nominal classifiers, because they are the most frequently used category, possessing the largest paradigm.

A nominal classifier “must occur with a number and/ or a demonstrative, or certain quantifiers before a noun” (Killingley, 1981, p. 104), with a structure of “Q (quantifiers) + C (classifiers) + N (head noun)”, as shown in the following two examples:

Example 4-4
一 只 猫
yi  zhi  mao
one  CL(animate-inhuman)  cat
one cat

Example 4-5
一 袋 米
yi  dai  mi
one  CL (measuring unit)  rice
one bag of rice

In fact, examples 4-4 and 4-5 represent two types of nominal classifiers: nominal classifiers in the first type “categorize the noun’s referent by identifying some salient perceived or imputed feature of the head noun, thus qualifying the head noun”, and only “co-occur with certain classes of nouns,” such as the classifier zhi in Example 4-4; those belonging to the second type are “quantifying the head noun,” and indicate “a temporary measuring unit” (Hu, 1993, p.9), such as the classifier dai in Example 4-5.
The first type is known as a qualifying classifier, while the second type is called a quantifying classifier.

There are overwhelmingly more quantifying classifiers than qualifying classifiers. The estimated number of classifiers runs from as many as 900 in *Hanyu Liangci Cidian* 汉语量词词典, “A dictionary of Chinese classifiers” (Bitchener, 1999), to as few as several dozens (Chao, 1968; Li & Thompson, 1981). However, qualifying classifiers were estimated at only about 70 (Hu, 1993).

Although the number of qualifying classifiers is much smaller than that of quantifying classifiers, the acquisition of qualifying classifiers challenges L1 Chinese children more.

Qualifying classifiers include a general classifier and dozens of specific classifiers. Specific qualifying classifiers categorize nouns into semantic classes based on their different physical properties, such as shape, size, animacy, etc. (Harred et al., 1972; Hu, 1993), making the classifier selection in Chinese an extremely complex task for young children. Children have to choose appropriate classifiers by relying solely on the semantic co-occurrence constraints on the classifier–head noun structure, with no morphological cues.

The most frequently used qualifying classifier in Mandarin is 个 *ge*, also known as a general classifier. It is usually the first classifier acquired by Chinese children (Erbaugh, 1986; Hu, 1993; Liu, 2008; Loke, 1991; Loke & Harrison, 1986; Lu & Li, 2008). Specific classifiers emerge after the general classifier, following a certain order determined by the perceptual salience of the object properties. For instance, children
were found to acquire shape classifiers earlier than function classifiers (Loke & Harrison, 1986).

After the above brief sketch of Chinese classifiers, I will now discuss the theoretical consideration for choosing Mandarin qualifying classifiers as the target linguistic structure in the current study. First of all, although a conventional association exists between a qualifying classifier and a fixed set of head nouns, its communicative value is quite limited, as the general classifier \(ge\) could almost always be used to replace the specific classifier within a context. Doing so usually will not jeopardize the comprehension of the utterance. Secondly, unlike gender agreements in the Romance languages, no morphophonological inflection occurs between the classifiers and the head noun. The classifier is a single, discrete item to be learned, with no further nominations on the sentence. Thus the feedback given in the treatments could provide learners with a context in which a classifier should be applied, and enable them to (re)acquire it within a relatively short period of time.

From a methodological aspect, classifiers are relatively easy to elicit in meaning-focused interaction. As a classifier always associates with a head noun, in the two communication tasks in the current study, head nouns were used as stimuli to elicit the target classifiers.

Lastly, on pedagogical grounds, previous studies showed that Chinese L1 children usually do not complete their acquisition of classifiers until preschool or early school years (Chang, 1983; Erbaugh, 1982; Liu, 2008; Lu & Li, 2008; Mak, 1991). Classifiers start to emerge in the children’s utterances as early as 1; 07 (Szeto, 1998). By three years old, they acquire the general classifier \(ge\) 个, but tend to use it to replace
any specific classifiers (Erbaugh, 1982; Hu, 1993). This overgeneralization of *ge*
continues until they are four to five years old (Liu, 2008; Lu & Li, 2008). By six years old,
children rapidly improve their ability to use appropriate specific classifiers (Erbaugh,
1982; Liu, 2008). However, they do not fully master the classifier system until the age of
nine (Mak, 1991). Acquiring classifiers is also a challenge for learners of Chinese as a
FL, especially when their L1 is a non-classifier language, such as English.

**The Mandarin classifiers in the current study.** In the current study, 20 Mandarin
qualifying classifiers (see Table 4-1) were chosen from *Integrated Chinese I and II*;
*these classifiers* were introduced as either required or supplementary vocabulary in the
classroom. These classifiers belong to five categories: animacy, shape, arrangement,
function, and the general classifier *ge*. In the treatments, 29 nouns were given as stimuli
to elicit the target classifiers. Among them, 11 classifiers used two stimuli. Since 双
shuang shared the same two stimuli with 只 zhi, one more stimulus was added. Seven
other specific classifiers have a more limited choice of head nouns, and therefore only
one stimulus was used. The general classifier could be used to replace any of the
specific classifiers. The only exception is 枝 zhi, when used to modify a flower with one
or several blossoms on the stem.

4.2.5 *Materials*

4.2.5.1 **Background information questionnaire**

The background information questionnaire (see Appendix C) was designed to
achieve the following two goals: first, identifying the two types of targeted population in
the current study: CHL learners who grew up in a Mandarin-speaking family, and non-
HL learners with English as their L1. Second, collecting information regarding HL
learners’ language experience and home environment in order to help the researcher to prepare the interview guide (see Section 4.2.5.5). The questionnaire consisted of 14 items, 10 of which were closed-item questions that related to the most fundamental information about the learners’ language experience, such as learners’ birthplace, L1, home language, etc.

Table 4-1. The targeted classifiers.

<table>
<thead>
<tr>
<th>Semantic Domain</th>
<th>Classifier</th>
<th>Meaning</th>
<th>Stimuli Nouns</th>
<th>Stimuli Nouns</th>
</tr>
</thead>
<tbody>
<tr>
<td>Animacy</td>
<td>只 zhi</td>
<td>inhuman</td>
<td>鸡, 猫</td>
<td>chicken, cat</td>
</tr>
<tr>
<td></td>
<td>匹 pi</td>
<td>inhuman</td>
<td>马</td>
<td>horse</td>
</tr>
<tr>
<td></td>
<td>条 tiao</td>
<td>inhuman</td>
<td>龙, 蛇</td>
<td>dragon, snake</td>
</tr>
<tr>
<td>Arrangement</td>
<td>双 shuang</td>
<td>paired-item</td>
<td>筷子, 鞋, 手套</td>
<td>chopsticks, shoe,</td>
</tr>
<tr>
<td></td>
<td>只 zhi</td>
<td>single-item</td>
<td>手套, 鞋</td>
<td>gloves, shoe</td>
</tr>
<tr>
<td>Function</td>
<td>把 ba</td>
<td>hand-tool</td>
<td>牙刷, 剪刀</td>
<td>toothbrush, scissors</td>
</tr>
<tr>
<td></td>
<td>件 jian</td>
<td>upper-body garments</td>
<td>衬衫, 毛衣</td>
<td>shirt, sweater</td>
</tr>
<tr>
<td></td>
<td>顶 ding</td>
<td>supported objects</td>
<td>帽子</td>
<td>hat</td>
</tr>
<tr>
<td></td>
<td>架 jia</td>
<td>vehicles</td>
<td>钢琴</td>
<td>piano</td>
</tr>
<tr>
<td></td>
<td>辆 liang</td>
<td></td>
<td>汽车</td>
<td>car</td>
</tr>
<tr>
<td>Shape</td>
<td>条 tiao</td>
<td>long-flexible</td>
<td>领带, 腰带</td>
<td>tie, belt</td>
</tr>
<tr>
<td></td>
<td>朵 duo</td>
<td>drooping flowers</td>
<td>花</td>
<td>flower</td>
</tr>
<tr>
<td></td>
<td>枝 zhi</td>
<td>long, hard, stick-like</td>
<td>笔, 花</td>
<td>pen, flower</td>
</tr>
<tr>
<td></td>
<td>张 zhang</td>
<td>flat-thin</td>
<td>画, 桌子</td>
<td>picture, desk</td>
</tr>
<tr>
<td></td>
<td>面 mian</td>
<td>round-small</td>
<td>镜子</td>
<td>mirror</td>
</tr>
<tr>
<td></td>
<td>颗 ke</td>
<td>block-like</td>
<td>珍珠, 葡萄</td>
<td>pearl, grape</td>
</tr>
<tr>
<td></td>
<td>块 kuai</td>
<td>flat and thin</td>
<td>牛肉, 月饼</td>
<td>beef, mooncake</td>
</tr>
<tr>
<td></td>
<td>片 pian</td>
<td>needle</td>
<td>树叶, 面包</td>
<td>leaf, breadcake</td>
</tr>
<tr>
<td></td>
<td>根 gen</td>
<td></td>
<td>针</td>
<td>needle</td>
</tr>
<tr>
<td>General</td>
<td>个 Ge</td>
<td>General</td>
<td>general</td>
<td>general</td>
</tr>
</tbody>
</table>
4.2.5.2 Treatment sessions

In designing the treatment tasks, the following issues were carefully taken into consideration: task effects and contextual factors.

Previous studies suggested that learner factors (e.g., ethnicity, gender, proficiency, etc.) affected their learning in task-based instructional settings (Bitchener, 1999; Gass & Varonis, 1986; Slimani-Rolls, 2005). Given the heterogeneous nature of the participants in the current study, each task was carefully designed to make sure it not only elicited the targeted classifiers, but also enabled each learner to complete it at their most comfortable level. In particular, HL learners usually have greatly advanced speaking skills compared with their non-HL counterparts (Valdés, 1995; 2001) even though they are enrolled in the same class, and present similar or even weaker grammatical knowledge on standard tests. Therefore the pictures used for stimuli were carefully chosen (such as the color and shape of items, the actions of animals, etc.) to make sure that learners had a lot to say when their language abilities allowed. Thus more interactions could occur.

According to the flow of the information, tasks that are commonly employed in SLA research can be divided into two categories: one-way tasks and two-way tasks (Doughty & Pica, 1986; Long, 1988; Pica, 1987; Varonis & Gass, 1985). In a one-way task, only one party of the dyad holds the information that is needed to complete the task. Thus during the interaction, the information flows from the information holder to the other party of the dyad. In contrast, in a two-way task, both parties of the dyad have the information, and they are expected to complete the task through exchanging the information. Previous studies have provided controversial results in terms of the task effects of one-way and two-way tasks: many researchers have argued that two-way
tasks generate more negotiations of meaning among learners (Doughty & Pica, 1986; Gass & Varonis, 1985; Long, 1988; Pica, 1994), but others have questioned this argument (Duff, 1986; Nakahama et al., 2001). In order to minimize the task effects, the current study employed both a one-way task and a two-way task.

Lastly, the context in which feedback was provided was designed to enable recasts to serve multiple discourse functions, in order to make it as implicit as possible. Previous research showed that recasts could be more or less implicit, depending on various recast features (e.g., length, number of changes, prosodic emphasis, etc.) and the discourse context (Loewen & Philp, 2006). For instance, in the classroom, learners are sometimes not sure whether a recast provided by the teacher was a corrective feedback or an approval of their response, since teachers often use both. In the tasks designed for the current study, although recasts were only provided on errors made by learners, they were done in as natural a way as possible. The learners could therefore interpret them simply as a discourse move made by the researcher to carry on the interaction. I will be looking at precisely this question, specifically seeking to determine whether CHL and non-CHL learners perceived this feedback differently.

With the above careful considerations, the current study employed two task-based NS-NNS conversational interactions in the treatment sessions. In each interaction session, learners completed one communicative task. On Day 1, they engaged in a one-way interaction task (“a story telling”); on Day 2, they completed a two-way task (“spot the differences”). During the interaction, participants received feedback on not only the problematic classifiers, but also any errors that they made. Both tasks used Power Point slides to elicit the target form. On each slide, several pictures were
displayed. Each picture portrayed the items that reflected the target head nouns with one picture usually portraying one target item (see Appendix F).

In the one-way interaction task, learners were given a set of Power Point slides, on each of which four to nine target items (e.g., furniture, food, an animal, etc) were displayed. Learners were required to tell a story using all of the provided items. In their story, they needed to provide the name and the quantity of each item. The interlocutor asked questions in Chinese (e.g., what did you see at the zoo? What did you buy for your mother for Mother’s Day?) to help learners develop their story.

In the two-way interaction task, each member of the dyad was shown a set of Power Point slides on two separate computer screens. Care was taken to make sure that neither member of the dyad could see the screen of the other party. The goal of the task was to find the differences between the learner’s and the researcher’s Power Point slides, relying solely on the conversational interaction. Each slide contained 3-6 items. The same items were already used to elicit classifiers in the one-way interaction task, but using different pictures in the learners’ set of Power Point slides.

4.2.5.3 Testing material

In previous experimental studies, various tests were employed to measure the effects of feedback. Some of them favored explicit knowledge, such as grammaticality judgment (Muranoi, 2000; Nagata, 1993), sentence completion (e.g., Sanz, 2003), and translation tests (e.g., Havranek & Cesnik, 2001). Others assessed learners’ implicit knowledge, such as picture description (DeKeyser, 1993; Leeman, 2003), and story-telling tasks (DeKeyser, 1993). Tests designed to measure explicit knowledge provide learners with unlimited time, which enables them to have enough time to apply their
meta-linguistic knowledge. On the other hand, tests designed to measure implicit knowledge expect learners to produce the TL in a spontaneous manner within a limited time period.

However, solely relying on one type of tests harmed the construct validity of the research results. For instance, R. Ellis (2007) criticized some research on the effects of feedback for relying too much on testing instruments that better measure explicit L2 knowledge. As a result, explicit feedback might be found to be more effective than implicit feedback in these studies. In order to avoid the above problems, it was decided to employ multiple testing instruments that could measure both implicit and explicit knowledge, as a solution to fully portray the degree of improvement caused by the feedback.

R. Ellis and his colleagues (R. Ellis, 2007; R. Ellis et al., 2006) used three different types of tests to measure the effects of implicit and explicit feedback on English past tense -ed: an oral imitation test, an untimed grammaticality judgment test, and a meta-linguistic knowledge test. Among the three tests, the oral imitation test aimed at measuring learners' implicit knowledge, while the un-timed grammaticality judgment test and the meta-linguistic knowledge test were used to assess learners' explicit knowledge. The oral imitation test consisted of 36 audio-recorded belief statements. During the test, participants first indicated on an answer sheet whether they agreed, disagreed, or were not sure about the statement. Next they were requested to repeat each statement in correct English. The untimed grammaticality judgment test consisted of 45 written sentences, including seven grammatically correct sentences and eight grammatically incorrect sentences, as well as 30 distractor sentences. Learners were
required to indicate whether each sentence was grammatically correct or not, report
their degree of certainty, and indicate whether they applied any rules in their judging
process. The meta-linguistic knowledge test was formed by five sentences. During the
test, learners were told that these sentences were ungrammatical, and were requested
to correct the error, and provide the reason for the errors. The results showed significant
group differences in the oral imitation test, but not in the untimed grammaticality
judgment test in terms of ungrammatical sentences. No significant group differences
were found in both tests in terms of the grammatical sentences.

Loewen and Nabei (2007) used three different testing instruments in their
experiment that investigated the effects of feedback on English question formation: a
timed grammaticality judgment test that limited the completion time of each item to 1.8
to 5 seconds, an un-timed grammaticality judgment test, and an oral production task.
The untimed grammaticality judgment test was designed to measure learners’ L2
explicit knowledge, while the oral production task assessed their L2 implicit knowledge.
The same set of 40 items was used in both the timed and untimed grammaticality
judgment tests. Thus any differences in learners’ performances could be interpreted as
differences in their two types of knowledge: when the time was controlled, learners were
pushed to use their “feel,” and there was “little need or opportunity to access meta-
linguistic knowledge” (R. Ellis, 2005, p.157). On the other hand, the untimed task
allowed learners to involve their meta-linguistic knowledge. The oral production task
consisted of two-spot-the-differences tasks. The results showed an increase in the
timed grammaticality judgment test, but not in the untimed grammaticality judgment test
on the oral production test.
In sum, the above studies showed that employing tests designed to tap different types of linguistic knowledge generated mixed results, which provided deeper insights for researchers to investigate the effects of implicit and explicit feedback.

Based on the review of these previous studies, the current study employed two types of instruments: an oral imitation test and a written cloze test. On the oral imitation test, learners would be focused primarily on meaning under pressure to perform in real time\(^6\), therefore the test was predicted to measure learners’ knowledge that is more intuitive and implicit. On the other hand, learners on the untimed written test were predicted to be more focused on form, therefore it was expected that this task would measure learners’ knowledge that is more rule-based and explicit. Their literacy skills were particularly important for completing this task.

4.2.5.4 The oral imitation test

The oral imitation test consisted of 33 audio-recorded items. Among them, 22 items related to the targeted classifiers. Half of them used correct classifiers, and the other half used incorrect ones. Moreover, 11 items served as distractors. All 33 items were in the form of declarative sentences.

In designing the testing items, care was taken to make sure of the following issues: first, all the vocabulary used in the testing items was chosen from *Integrated Chinese I and II*. In other words, all the participants in the current study should have already acquired, or at least encountered these vocabulary words at the time the experiment was conducted. Second, each sentence was made long enough (containing approximately 10-14 words) to make sure that it exceeded learners’ STM, so that they needed to tap their implicit knowledge rather than simply memorizing it in order to complete the task (Mackey & Gass, 2005).
During the test, the audio-recording of the test items was played one item at a time. Each item was played only once. The learners were required to first indicate orally whether they agreed with, disagreed with, or were unsure about the statement in each item based on what they heard. Learners were told that they could disagree when they spotted anything wrong in the sentence (e.g., grammar, choice of vocabulary, etc). However, regardless of which indication they gave, learners were required to repeat each sentence. If they disagreed with the sentence, they were asked to repeat the sentence in correct Chinese, as shown in Example 4-6.

Example 4-6
我昨天去商店买了三条筷子
*Wo zuotian qu shangdian mai le san tiao kuaizi.*
1sg yesterday go shop buy PFV three CL chopsticks
Yesterday I went to shop and bought three chopsticks.

The learner was expected to disagree with the sentence, since the classifier used in the sentence was not correct. Upon the completion of the judgment, the learner was requested to repeat the sentence orally in “correct” Chinese. “The basic assumption underlying elicited imitations is that if a given sentence is part of one’s grammar, it will be relatively easy to repeat” (Mackey & Gass, 2005, p.55). Classifiers generally have relatively low communicative value in a Chinese conversation, since they can almost always be replaced with the general classifier ‘ge’; learners would have not focused their attention on classifiers if they were not already in their lexicon, especially when the sentence is reasonably long. Other elements in the sentence, such as verbs, nouns, or adjectives might more likely draw their attention compared with classifiers.

Although no minimal time was set to complete the oral imitation test, participants are still under time pressure as each of the 33 audio-recorded items was only played once.
As soon as the participant produces his/her utterance, the next item will be played. There is not time for reviewing or checking back.

4.2.5.5 The untimed written cloze test

The written test consisted of 22 items, each of which was in the form of a declarative sentence. As with the oral test, all the vocabulary in the written test appeared in *Integrated Chinese I and II*. Therefore participants should have already acquired or at least encountered these items in the classroom. During the test, learners were requested to fill in the blanks in Chinese characters by referring to the matched English sentence, as shown in example 4-7. They were told to write in Hanyu Pinyin when they could not remember the Chinese characters.

Example 4-7

我 有 三 只 猫 和 两 匹 马。

Wo  you  san  zhi  mao  he  liang  pi  ma  
1sg  have  three  CL  cat  and  two  CL  horse  
I have three cats and two horses.

In example 4-7, the learner was expected to provide the two missing short phrases 三只猫 and 两匹马, based on the matched English sentence “three cats” and “two horses”. Each phrase requested a structure of “number + classifier + head noun”. There was no time limitation set for completion of the test.

Besides the 21 targeted classifiers, 12 other words were also taken out of same sentences to serve as distractors, as shown in example 4-8.

In example 4-8, besides the phrase 一 双 鞋, which contained the targeted classifier 双, learners were also requested to provide the Chinese equivalent of the word “birthday”.

129
4.2.5.6 **Advantages and disadvantages of the testing design in the current study**

The pre-test-post-test design enabled me to directly examine the effects of feedback in the current study. In the SLA literature, the effects of feedback were also measured by uptake, which was defined as learners’ immediate response toward NS’s corrective feedback (e.g., Loewen, 2004; Lyster, 1998b; Panova & Lyster, 2002). However, its validity was questioned by a number of researchers (Long, 2007; Lyster, 1998a; 2004; Ohta, 2000). Long argued that *uptake* episodes produced toward different types of feedback should not be compared with each other since different types of feedback were designed to achieve different instructional goals. For instance, explicit feedback (e.g., meta-linguistic, elicitation, repetition) usually elicits more uptake than implicit feedback (e.g., recast) does as it was designed to push learners to produce more output in the first place. In addition, the occurrence of learner uptake significantly relates to various characteristics of recasts, such as the length of recasts, the linguistic focus, the type of change, the number of changes, etc. (Sheen, 2006). Moreover even when learners do produce uptake following recasts, it does not necessarily lead to subsequent L2 development. In fact, learners with lower proficiency may not be able to identify the corrective intention embedded in recasts, especially when their
developmental level is not ready for the target forms. Their responses toward recasts could be just red herrings (Mackey & Philp, 1998). By employing a pre-test and post-test design, feedback under investigation could be manipulated in a context in which the independent variables were carefully controlled. In addition, as learners’ previous knowledge of the target forms could be measured, any improvement they gained during the post-test could be attributed to the effects of the treatments. Thus, the correlation found between the effects of feedback and learners’ language development was more reliable.

On the other hand, the pre-test-post-test design also has disadvantages. First of all, as the pre-test was administrated before the treatment session, it could cue learners to the target linguistic forms. Particularly, since the HL learners investigated in the current study had already stored a considerable amount of implicit knowledge of the target forms from their childhood HL exposure, the pre-test may help them “reactivate” that knowledge. In contrast, the non-HL learners have had exposure to these items only through formal classroom instruction. Therefore it may not be equally likely for them to “remember” the forms that they once learned but forgot later, since these forms were never fully mastered. In other words, the pre-test-post-test design could favor HL learners more than it does non-HL learners. Second, the pre-test-post-test design does not allow the researcher to measure the effects of any feedback that arises incidentally during the interaction (Nassaji, 2009).

4.2.5.7 Measurement of learners’ perception of the feedback

During the interaction, learners’ attention is usually directed toward particular aspects of the language. However, those aspects may not always be the aspects that
the NS intended to call learners’ attention to with the corrective feedback. Previous studies showed a correlation between learners’ noticing of the corrective intention of the feedback and L2 development (Adams, 1991; Egi, 2007; Mackey, 2006; Rosa & Leow, 2004; Swain & Lapkin, 2002).

In the literature, learners’ noticing of feedback has been examined mainly using introspective methods, in which learners “communicate their internal processing and perspectives about language learning experiences” (Mackey & Gass, 2005, p. 201) either in a written form, such as learning journals, diaries, questionnaires, etc, or in verbal reports, such as think-aloud protocols, prompted repetition of feedback, metatalk, stimulated recall, immediate recall, etc. These methods can be either conducted online (in real time), or off line (not in real time) (see Egi, 2004, for a review of these methods).

Among the above-mentioned methods, verbal reports have been employed as a standard research tool to examine learners’ noticing at the level of attention and awareness. Verbal report is generally defined as “a special type of introspection and consists of gathering protocols, or reports, by asking individuals to say what is going through their minds as they are solving a problem or completing a task” (Mackey & Gass, 2005, p.77). In order to be taken as evidence of noticing (Schmidt, 2001, p. 77), verbal reports have to be collected either concurrently or immediately following the experience. The two most commonly used types of verbal reports are stimulated recall (Adams, 1991; Mackey et al., 2000; Nabei & Swain, 2002), and immediate recall (Egi, 2004; Leow, 1997; 2000; Philp, 2003; Rosa & O’Neill, 1999).
In a stimulated recall, learners are prompted to recall thoughts they had while performing a task or participating in an event. Through the use of stimulated recall, a subject may be “enabled to relive an original situation with great vividness and accuracy if he/she is presented with a large number of cues or stimuli which occurred during the original situation” (Bloom, cited in Mackey & Gass, 2005, p.17).

Immediate recall elicits data “immediately after the completion of the event to be recalled” (Mackey & Gass, 2005, p.85). During the interaction, learners are asked to verbalize their thoughts or repeat the last thing of the recast they heard during a conversational turn, immediately after a recall prompt signal (Egi, 2004; Philp, 2003, p.85).

Both stimulated recall and immediate recall are important methods in eliciting noticing data of recasts and other corrective feedback. Both have their advantages. The biggest advantage of immediate recall over stimulated recall lies in the fact that it is “free from a memory decay problem” since it elicits noticing data immediately following the event when the information is still fresh in learners’ STM, thus it “captures the critical moment of noticing” (Egi, 2004, p.259). On the other hand, stimulated recall is superior to immediate recall in the following aspects: first, it doesn’t interrupt the communication flow since it is carried out after the completion of the oral interaction. Second, the researcher can clarify any uncertainties and doubts about learners’ comments on the spot. In addition, stimulated recall requires less training prior to data collection compared with immediate recall for both the researchers and the participants.

The current study chose stimulated recall over immediate recall due to the following considerations: first, the two communication tasks employed in the experiment
required a smooth communication flow with as little interruption as possible. Particularly during the story-telling task, if a learner is frequently interrupted by prompt signals, he/she may lose interest in carrying on the story. Second, in previous studies, the average length of immediate recall protocols was significantly shorter than stimulated recalls since the immediate recall was carried out under the pressure of continuing the interaction (Egi, 2004). As learners’ perception of feedback is one of the dependent variables investigated in the current study, the more detailed information provided by the learners, the more their perception of this feedback could be explored.

The procedure of the stimulated recall in the current study followed Gass and Mackey (2000). A typical example of stimulated recall is shown below:

Example 4-9
N: Ok, Let’s look at the video we took from yesterday.
   (Played the video.)
L: *我  买了    一件    帽子。
   Wo mai le    yi    jian    maozi.
   1sg    buy    PFV    one    CL    hat.
   I bought one hat.
N: 你  买了    一顶    帽子?
   Ni mai le    yi    ding    maozi.
   2sg    buy    PFV    one    CL    hat.
   You bought a hat?
L:  我  买了    一顶    帽子.
   Wo mai le    yi    ding    maozi
   1sg    buy    PFV    one    CL    hat.
   I bought one hat.
N:  (Paused the video.)
   What were you thinking at that time?
L:  I didn’t really know how to say one hat. I was not sure which measure word I should use for “hat”. And she (the NS) corrected me.

In the above example, the researcher first let the learners watch a video clip taken from the treatment in which the learner received the feedback. Then she paused the video, and asked the learner to verbally report what he had in mind when the feedback
episode occurred. Upon the completion of the learner’s recall, the researcher started to play the next video clip.

4.2.5.8 Interview

As discussed in Chapter 2, HL learners’ linguistic proficiency should not only be assessed by their visible language performance, but also by their underlying language competence, which developed over time through their language exposure (Hendryx, 2008). Thus, we should not rely solely on traditional tests to assess a HL learner’s proficiency level. In fact, the more we know about a learner’s past language experience and home language environment, the closer we come to describing his/her real linguistic knowledge and abilities (Jiang, 2008). The ideal approach to fully portray a HL learner’s existing linguistic system is to record every step of his/her language development through longitudinal research. Obviously, it was not realistic to do so when many participants were investigated, as in the current study. Thus the current study employed a semi-structured interview. There are several advantages to using this data collection method: first, the researcher could collect needed information by asking pre-prepared, well-organized questions. Secondly, the researcher could ask in-depth or unstructured questions when necessary, with sufficient flexibility. Thus, the researcher could collect all the intended information and get a big picture of all the HL participants, while simultaneously collecting specific information concerning each individual learner (Heigham & Sakui, 2009).

Previous studies have shown that HL learners’ proficiency can be predicted by a variety of social variables, such as age of arrival, cultural identity, socioeconomic status, language environment, etc. (Jia, 2008). Moreover, motivation was also found to play an
important role in CHL learners’ language achievement and confidence about their language skills (Lu & Li, 2008). In addition, home environment was found to have a significant impact on HL learners’ morphological awareness and literacy development (Koda et al., 2008a; Koda et al., 2008b; Xiao, 2008). Based on the findings of these previous studies, the interview in the current study consisted of four main parts: a. language background information about the learner (e.g., What was the first language or languages they spoke? At what age did English become their dominant language? Did they receive any systematic literacy education either at home or in Chinese-community schools? b. their motivation for learning Chinese; c. background information about their parents (e.g., What do their parents do for a living? What languages do they speak? How are their English skills in the traditional four domains?); d. the home environment (e.g., What are the available Chinese print materials at learners’ home? How often do parents and children engage in parent-child HL literacy-related home activities? ). Before each interview, the researcher prepared an interview guide based on careful reviews of the background questionnaire of each interviewee. The interview guide contained questions that both were general and individual-specific. During the interview session, the researcher mostly followed the interview guide. However, she also listened to the interviewee very carefully in order to ask flexible questions that were not pre-prepared, and thus completed each learner’s unique language profile.

4.2.6 Procedure

The experiment was conducted on three consecutive days. At the beginning of the experiment, all participants were requested to sign the consent form and fill out the background information questionnaire. Based on the self-reported information in the
questionnaire, the participants were identified as either HL learners or non-HL learners, and were randomly assigned to a group (see Table 4-2).

Table 4-2. The experimental and control groups.

<table>
<thead>
<tr>
<th>Recasts</th>
<th>Meta-linguistic</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>CHL 12</td>
<td>Non-CHL 12</td>
<td>CHL 11</td>
</tr>
</tbody>
</table>

The 47 learners in the experimental groups then took the pre-test. They first took the oral test, then the written test. Conducting the oral test before the written test was due to the following considerations: since written test was untimed, learners would have adequate time to tap their explicit knowledge. Therefore if they took the oral test after taking the written test, they could easily draw the explicit knowledge that was still fresh in their memory instead of tapping their implicit knowledge, which would defeat the purpose of the test design. By employing oral test before the written test, learners could be pushed to use their explicit knowledge to a greater extent in the written test if they encountered problems that could not be resolved with their implicit knowledge in the oral test. Following the pre-tests, they completed the picture description task in Treatment 1. On Day 2, The 47 learners in experimental groups first participated in the spot-the-differences task in Treatment 2, then took the post-test. Again they first took the oral test, then the written test. On Day 3, all the learners in experimental groups participated in the stimulated recall session, in which they met the researcher individually. After the stimulated recall, the non-HL learners exited the experiment. The 23 HL learners were interviewed by the researcher individually. The procedure and approximate time spent for each task are summarized in Figure 4-2.
The 17 Learners in the control group took the pre-test on Day 1. They also first took the oral test, then the written test. On day 2, they completed the oral and written post-tests in the same order as they did in pre-test.
4.2.7 Transcriptions

The researcher transcribed all the tests, treatments, stimulated recall, and the interview data. For the pre- and post-test data (128 tests, approximately 23 hours), only the classifier used by the learners was transcribed. For the treatment data (94 tasks, 7.45 hours for CHL learners, 11.9 hours for non-CHL learners, the stimulated recall data (6.35 hours for CHL, 14.2 hours for non-CHL learners), and the interview data (13 hours), the entire recording was transcribed.

4.2.8 Scoring

On oral tests, two points were awarded for target language (TL) production, and zero points for non target language (NTL) production. One point was given when a learner used _ge_ to replace the appropriate classifier. TL production was operationalized as TL suppliance of the targeted classifier (e.g., _jian_ for _chenshan_ "shirt"). NTL production was operationalized as NTL suppliance of the classifier (e.g., _tiao_ for _chenshan_ "shirt").

On written tests, three points were awarded for TL production in Chinese characters (e.g., _件_ (jian) for _chenshan_ “shirt”), two points for TL production in Hanyu Pinyin\(^7\) (e.g., _jian_ for _chenshan_ “shirt”). One point was awarded for the general classifier _ge_, regardless of whether in Chinese characters or in Hanyu Pinyin\(^8\). Zero points were given for NTL production (e.g., _tiao_ for _chenshan_ “shirt”).

4.2.9 Coding

The stimulated recall protocols were coded according to perception categories that were generated based on previous research on learners’ perception about the target of the feedback. Learners’ perception of feedback has been coded into six categories in previous research—lexical, semantic, phonological, morphosyntactic, no content, and
unclassifiable (Mackey et al., 2000; Gass & Lewis, 2007). Since the current study focused on classifiers, the learners’ recall comments were first divided into two categories based on whether the learner perceived the feedback as classifier-related. Therefore the learners’ recall comments were first divided into either “classifier-related comments” or “non classifier-related comments”.

“Classifier-related” comments were further coded into the following four subcategories: reacquisition (R), acquisition 1 (A1), acquisition 2 (A2), and other (O). “Non classifier-related” comments were further coded into the following five subcategories: Semantics (SE), phonology (PH), morphosyntax (MO), lexis (LE), and no content (NC). The coding scheme is summarized in figure 4-3.

![Figure 4-3. Coding scheme of the recall data](image)

### 4.2.9.1 Classifier-related comments

The four subcategories under classifier-related comments were tailor-made for the current study in order to examine learners’ previous knowledge of the target classifiers.
R: reacquisition was operationalized as learners’ comments that clearly indicate they knew which measure word should be used before participating in the study, as shown in Example 4-10.

Example 4-10

*Feedback episode*

NNS: *去动物园我看到了一个龙。

Qu dongwuyuan wo kan dao le yi ge long.
Go zoo 1sg see RC PFV one CL dragon
I went to the zoo and saw a dragon.

NS: 那条龙有多长？

Na tiao long you duo chang?
That CL dragon have how long
How long was that dragon?

NNS: 很长。

Hen chang.
Very long.
Very long.

*Recall data*

“我知道是“一条龙”，不过我忘记掉了。”
(I knew it should be ‘yi tiao long’, but I forgot.)

A1: Acquisition 1 was operationalized as learners’ comments indicating that they did not know which measure word should be used before participating in the study, and used the general measure word ‘ge’ to replace it, as shown in Example 4-11.

Example 4-11

*Feedback episode*

NNS: *我买了三个腰带。

Wo mai le san ge yaodai.
1sg buy PFV three CL belt.
I bought three belts.

NS: 你的三条腰带是什么颜色的？

Ni de san tiao yaodai shi shenme yanse de?
2sg NOM three CL belt is what color NOM
What color are your three belts?

NNS: 三条。一个 dao 是红色的，还有黑色的。

San tiao. Yi ge dao shi hongse de. Hai you heise de.
Three CL. One CL UK is red NOM Also have black NOM
Three. One is red, Also have (one) black.

*Recall data*

Again when I was doing that, I know I didn't know the measure words. It makes, kind of, my personality, I like to know what I talk about. I knew I just kept saying "ge" because I didn't know what the actual measure word was.

A2: Acquisition 2 was operationalized as comments made by learners to indicate that they thought a wrong measure word as the correct one, as shown in example 4-12.

Example 4-12  
*Feedback episode*

NS: 你看到多少蛇?

Ni kan dao duoshao she?  
2sg see RC how many snake  
How many snakes did you see?

NNS: *一 zhang 的蛇。

Yi zhang de she.  
One CL NOM snake  
One snake.

NS: 你看到一条蛇?

Ni kan dao yi tiao she?

2sg see RC one CL snake  
You saw one snake?

NNS: 一条

YI tiao.  
One CL.  
One (snake).

*Recall data*

The measure word again. I have learned snake before. Second year, but I, it was the first semester of advanced, and it was really difficult, I kind of remember, it was a story, like a fairy tale, but we never learned it. And I learned "tiao", but never for snake. I know it is for long object, but didn't know that was for long object, but we always used it for "he", or I mean really long and thin object. You never put it with animals. I thought it is for inanimate objects.

O: The learner didn't use any measure word in the original sentence as shown in example 4-13.
Example 4-13
*Feedback episode*

NNS: *我买四笔。

Wo mai si bi.
1sg buy four pen
I bought four pens.

NS: 你买了四枝笔?

Ni mai le si zhi bi.
2sg buy PFV four CL pen
You bought four pens.

NNS: 四枝笔。

Si zhi bi.
Four CL Pen
Four pens.

*Recall data*
Again with the measure word, it's always the measure words.

4.2.9.2 Non classifier-related comments

SE: the semantic category was operationalized as general comments about communicating meaning, creating understanding, or being unable to express an intended meaning.

Example 4-14
*Feedback episode*

NNS: 我住在这儿佛罗里达。那儿很大，很。。。

Wo zhu zai zher foluolida. Nar hen da, hen
1sg live in here Florida. There very big, very.
I live in Florida, There is very big.

NS: 很热。

Hen re.
Very hot.

NNS: *所以我买了一个帽子。

Suoyi wo mai le yi ge maozi.
So 1sg buy PFV one CL hat.
So I bought one hat.

NS: 你那顶帽子是什么颜色的？

Ni na ding maozi shi shenme yanse.
2sg that CL hat is what color
What color is your hat?
Recall data
I was actually trying to think very bright, like light.

PH: the phonological category was operationalized as specific comments about pronunciation.

Example 4-15
Feedback episode
NNS: *我买了一个花儿。
       Wo mai le yi ge huar.
       1sg buy PFV one CL flower
       I bought one flower.
NS: 只有一枝？
       Zhi you yi zhi.
       Only have one CL.
       Only one (flower)?
NNS: 一枝。
       Yi zhi.
       One CL.
       One (flower).

Recall data
I didn't realize the tone difference between painting and flowers, huar (T1) and huar (T2).

MO: the morphosyntax category was operationalized as comments about sentence formation and structure or word order, as well as comments on specific aspects of grammar.

Example 4-16
Feedback episode
NNS: *一个镜子。
       Yi ge jingzi.
       One CL mirror.
       One mirror.
NS: 你那面镜子花了多少钱？
    Ni na mian jingzi hua le duoshao qian
2sg that CL mirror spend PFV how much money
How much money did you spend for that mirror?

NNS: 面？
    Mian
    CL
    Mian

*Recall data*
I always forget the pattern you state the object that is something next to
you first, like English is next to that table, but in Chinese, you say table
next to is something.

LE: the lexical category was operationalized as containing specific comments
about a known or unknown word, including the provision of a synonym and comments
about a synonym, or the word itself.

Example 4-17
*Feedback episode*

NNS: *两个 yuegao？*
    Liang ge yuegao
    Two CL yuegao
    Two moon cakes.

NS: 两块月饼。
    Liang kuai yuebing
    Two CL moon cakes
    Two moon cakes.

NNS: 月饼。
    Yuebing.
    Mooncake.

*Recall data*
I just can’t remember "bing", I just know "gao".

NC: *no content* was operationalized as instances in which the subject
participated verbally in the recall, yet said nothing about the content.

Example 14-18
*Feedback episode*

NS: 还有什么？
    Hai you shenme?
Also have what?
What else (do you) have?

NNS: *还有一钢琴。
    Hai you yi gangqin.
    Also have one piano
    Also have one piano.

NS: 啊，你还买了一架钢琴。
    A, ni hai mai le yi jia gangqin.
    You also bought one piano.

你那架钢琴多少钱？
Ni na jia gangqin duoshao qian
    2sg that CL piano how much money
    How much money did you spend on that piano?

NNS: 很多，很多钱。
    Hen duo, hen duo qian
    Very many, very many money
    A lot of money.

*Recall data*
"Gangqin henduo qian". They are really expensive.

4.2.10 Inter-Rater Reliability

The researcher scored 100% of the test and coded 100% of the stimulated recall
data. A female NS Chinese graduate student assistant scored 10% of randomly-
selected test data and 10% of randomly-selected stimulated recall data. The student
assistant received a training session from the researcher, in which the scoring and
coding protocols were explained. In addition, a set of test and stimulated recall data
were also provided for the student assistant to practice on before beginning the formal
scoring and coding, the results of which were not included in the analysis.

A simple agreement percentage was employed to assess the consistency in
scoring and coding. This measure indicated high rates of agreement between the raters
for both data sets: inter-rater reliability was 99% for the test, and 95% for the stimulated
recall data.
Sheen (2006) found that declarative recasts elicited a higher amount of uptake than interrogative recasts.

The term “measure word” is used interchangeably with “classifier” in the current study.

Fujian belongs to Min dialect group, Wenzhou and Shanghai belong to Wu dialect group.

Szeto (1998) argued that ge 个/個 should only be treated as a candidate for the general classifier, but has not treated as the statues of it yet.

Integrated Chinese I and II are required textbooks in the first- and second-year Chinese program at UF.

Although no minimal time was set for the oral imitation test, learners were still under pressure to perform within limited time given audio-recorded sentence was only played once to the learner.

Hanyu Pinyin refers to the Romanization system for Mandarin.

Since the general classifier ‘ge’ is frequently used by learners to replace the appropriate specific classifier, no extra point has given when ‘ge’ was written in Chinese character.
This chapter presents descriptive and inferential statistics results for each research question. Research question 1 was about learners’ perception of recasts and meta-linguistic feedback. Research question 2 looked into the effectiveness of recasts and meta-linguistic feedback on CHL and non-CHL learners’ language development. Research question 3 addressed the relationship between feedback type, learners’ language background, and the degree of improvement on the post-test. All statistical analyses were conducted using SPSS Version 17.0 for Windows.

5.1 Research Question 1

Research question 1 asked whether CHL learners perceive both implicit and explicit feedback more accurately than non-CHL learners.

5.1.1 Descriptive Statistics

The database included 1,192 feedback episodes: 428 were provided for CHL learners, whereas 764 were for non-CHL learners. Of the 428 feedback episodes that were provided for CHL learners, 224 were recasts, whereas 204 were meta-linguistic feedback. On the other hand, of the 764 feedback episodes that were provided for non-CHL learners, 427 were recasts, 337 were for meta-linguistic feedback. Tables 5-1 and 5-2 and figures 5-1 and 5-2 present these results.

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>M</th>
<th>SD</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>RHL</td>
<td>224</td>
<td>18.67</td>
<td>7.79</td>
<td>52.34</td>
</tr>
<tr>
<td>MHL</td>
<td>204</td>
<td>18.55</td>
<td>11.01</td>
<td>47.66</td>
</tr>
</tbody>
</table>

Note. N = raw frequency; M = mean frequency of feedback episodes provided per learner in each feedback group; SD = standard deviation (frequency). MHL = CHL meta-linguistic feedback group; RHL = CHL recast group.
Table 5-2. Feedback episodes provided for non-CHL learner group

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>M</th>
<th>SD</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>RNHL</td>
<td>427</td>
<td>35.58</td>
<td>12.81</td>
<td>55.89</td>
</tr>
<tr>
<td>MNHL</td>
<td>337</td>
<td>28.08</td>
<td>5.78</td>
<td>44.11</td>
</tr>
</tbody>
</table>

Note. N= raw frequency; M = mean frequency; SD = standard deviation (frequency).

MNHL= non-CHL meta-linguistic group; RNHL= non-CHL recast group.

Of the 428 feedback episodes that were provided for CHL learners, 93.93% (n=402) of them were provided with introspective comments by learners. Of the 224 recast episodes, 96.88% (n=217) were commented on by learners. Similarly, of the 204 meta-linguistic feedback episodes, 90.69% (n=185) were commented on by learners. Table 5-3 summarizes these results. All these introspective comment data were collected during the stimulated recall.
<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>M</th>
<th>SD</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>RHL</td>
<td>217</td>
<td>18.00</td>
<td>7.79</td>
<td>96.88</td>
</tr>
<tr>
<td>MHL</td>
<td>185</td>
<td>18.00</td>
<td>10.86</td>
<td>90.69</td>
</tr>
</tbody>
</table>

Note. N = raw frequency; M = mean frequency; SD = standard deviation (frequency).

On the other hand, of the 764 episodes that were provided for non-CHL learners, only 91.75% (n=701) of them were commented on by learners. This breaks down as follows: of the 427 recast episodes, 88.76% (n=379) were provided with introspective comments; whereas of the 337 meta-linguistic feedback episodes, 95.55% (n=322) were commented on by learners. Table 5-4 presents these results.

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>M</th>
<th>SD</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>RNHL</td>
<td>379</td>
<td>33.25</td>
<td>12.26</td>
<td>88.76</td>
</tr>
<tr>
<td>MNHL</td>
<td>322</td>
<td>26.83</td>
<td>6.45</td>
<td>95.55</td>
</tr>
</tbody>
</table>

Note. N = raw frequency; M = mean frequency; SD = standard deviation (frequency).

In summary, learners across all the experimental groups provided introspective comments for more than eighty-eight percent of the feedback episodes. In the recast group, the percentage of the commented feedback episodes provided by the CHL learners (96.88%) was higher than those provided by the non-CHL learners (88.76%); in the meta-linguistic feedback group, the percentage of the commented feedback episodes provided by the CHL learners (90.69%) was slightly lower than those provided by the non-CHL learners (95.55%).

5.1.1.1 Classifier-related comments

Next I will report the results of classifier-related comments made by all the learners. Tables 5-5 and 5-6 summarizes the number, minimum and maximum, mean, standard deviations, and percentage of classifier-related comments provided by learners across the four experimental groups. For the CHL learner group, 71.43% of
introspective comments made by learners in the recast group were classifier-related, whereas the percentage made by learners in the meta-linguistic group was 89.19%. On the other hand, for the non-CHL learners, 57.26% of introspective comments made by learners in the recast group were classifier-related, whereas the percentage made by learners in the meta-linguistic group was 81.99%. Figure 5-3 graphically compares the percentage of classifier-related comments provided by learners across the four experimental groups.

Table 5-5. Classifier-related comments provided by CHL learner group

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>Minimum</th>
<th>Maximum</th>
<th>M</th>
<th>SD</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>RHL</td>
<td>155</td>
<td>24</td>
<td>91</td>
<td>69.35</td>
<td>22.5</td>
<td>71.43</td>
</tr>
<tr>
<td>MHL</td>
<td>165</td>
<td>71</td>
<td>100</td>
<td>91.31</td>
<td>9.58</td>
<td>89.19</td>
</tr>
</tbody>
</table>

Table 5-6. Classifier-related comments provided by non-CHL learner group

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>Minimum</th>
<th>Maximum</th>
<th>M</th>
<th>SD</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>RNHL</td>
<td>217</td>
<td>29</td>
<td>93</td>
<td>59.61</td>
<td>20.26</td>
<td>57.26</td>
</tr>
<tr>
<td>MNHL</td>
<td>264</td>
<td>62</td>
<td>94</td>
<td>81.91</td>
<td>11.29</td>
<td>81.99</td>
</tr>
</tbody>
</table>

Figure 5-3. Percentage of classifier-related comments across experimental groups

In summary, the descriptive statistics showed that learners in the meta-linguistic feedback group made more classifier-related comments than those in the recast group.
regardless of their language background. CHL learners made more classifier-related comments in both recast and meta-linguistic feedback groups compared with non-CHL learners in the recast group. However, CHL learners in the recast group did not outperform non-CHL learners in the meta-linguistic feedback group.

5.1.1.2 Non classifier-related comments

Tables 5-7 and 5-8 summarize learners' perceptions of feedback when they failed to make classifier-related comments.

Among the five sub-categories, the percentage of no content was over 50% for all the learners regardless of their experimental groups. This indicated that all the learners tended to give irrelevant comments on feedback when they did not make comments on the targeted Chinese classifiers. Following no content, lexis was the second most common category of comments made by all the learners.

However, there were also minor differences between the CHL learners and the non-CHL learners. In the recast condition, the third most common category was morphosyntax for the CHL learners, but phonology for the non-CHL learners. The least common category was phonology for the CHL learners, but semantics for the non-CHL learners.

In the meta-linguistic feedback condition, the CHL learners did not comment on any phonological and morphosyntactic issues, but 5% of their comments were about semantics. In contrast, semantics was the least common category of comments for the non-CHL learners, at 0%. The non-CHL learners also made limited comments on phonology (3.4%) and morphosyntax (1.7%).
Table 5-7. CHL learners’ perceptions of feedback not identified as classifier-related

<table>
<thead>
<tr>
<th>Group</th>
<th>PH (%)</th>
<th>SE (%)</th>
<th>MO (%)</th>
<th>LE (%)</th>
<th>NC (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>RHL (n=62)</td>
<td>1 (1.6)</td>
<td>3 (4.8)</td>
<td>5 (8.1)</td>
<td>18 (29.0)</td>
<td>35 (56.5)</td>
</tr>
<tr>
<td>MHL (n=20)</td>
<td>0 (0.0)</td>
<td>1 (5.0)</td>
<td>0 (0.0)</td>
<td>2 (10.0)</td>
<td>17 (85.0)</td>
</tr>
</tbody>
</table>

Note: PH=phonology, SE=semantics, MO=morphosyntax, LE=lexical, NC=No content

Table 5-8. Non-CHL learners’ perceptions of feedback not identified as classifier-related

<table>
<thead>
<tr>
<th>Group</th>
<th>PH (%)</th>
<th>SE (%)</th>
<th>MO (%)</th>
<th>LE (%)</th>
<th>NC (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>RNHL (n=162)</td>
<td>8 (4.9)</td>
<td>5 (3.1)</td>
<td>6 (3.7)</td>
<td>58 (35.8)</td>
<td>85 (52.5)</td>
</tr>
<tr>
<td>MNHL (n=58)</td>
<td>2 (3.4)</td>
<td>0 (0)</td>
<td>1 (1.7)</td>
<td>13 (22.4)</td>
<td>42 (72.4)</td>
</tr>
</tbody>
</table>

In summary, when learners failed to comment on classifiers, they were most likely to comment on things unrelated to the feedback episode, or to comment on lexical issues. There were minor differences between the CHL learners and the non-CHL learners in their comments on semantic, morphosyntactic, and phonological issues.

5.1.2 Inferential Statistics

In order to test whether CHL learners perceive recast and meta-linguistic feedback on Chinese classifiers more accurately than non-CHL learners, a one-way analysis of variance (ANOVA) was used to determine whether the means of classifier-related comments provided by learners significantly differed across all the experimental groups.

For the one-way ANOVA, the alpha level was set at .05. The results revealed a significant difference between the groups, $F (3, 43) = 7.77$, $p = .00$.

Since a significant between-group difference was found in learners’ provision of classifier-related comments, the null hypothesis of no differences between the groups was rejected. An LSD post hoc test was employed to locate the source of statistical significance. The results are summarized in table 5-9.

The LSD post hoc test revealed that CHL learners receiving meta-linguistic feedback provided significantly more classifier-related comments than CHL learners in the recast group ($p = .003$). They also significantly outperformed non-CHL learners in the
recast group \( (p = .000) \), but not those in the meta-linguistic groups \( (p = .192) \). One the other hand, CHL learners in the recast group did not significantly outperform non-CHL learners in either experimental group. These results suggested a superiority of meta-linguistic feedback over recast in terms of CHL learners’ perception of classifiers. In addition, it also indicated an advantage of CHL learners in the meta-linguistic feedback group over non-CHL learners in the recast group in terms of their accuracy of perceiving classifiers.

Table 5-9. Summary of results of LSD post hoc test

<table>
<thead>
<tr>
<th>Commented feedback</th>
<th>Group comparison</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>MHL&gt;RHL</td>
<td>.003*</td>
<td></td>
</tr>
<tr>
<td>RHL=RNHL</td>
<td>.167</td>
<td></td>
</tr>
<tr>
<td>RHL=MNHL</td>
<td>.077</td>
<td></td>
</tr>
<tr>
<td>MHL&gt;RNHL</td>
<td>.000**</td>
<td></td>
</tr>
<tr>
<td>MNHL&gt;RNHL</td>
<td>.002**</td>
<td></td>
</tr>
<tr>
<td>MHL=MNHL</td>
<td>.192</td>
<td></td>
</tr>
</tbody>
</table>

\* \( P < .05 \)  
\** \( P < .01 \)

Note. MHL = CHL meta-linguistic feedback group; RHL = CHL recast feedback group; RNHL = non-CHL recast feedback group; MNHL = non-CHL meta-linguistic feedback group

5.2 Research Question 2

Research question 2 asked whether explicit feedback was more effective than implicit feedback for CHL and non-CHL learners’ language development.

5.2.1 Descriptive Statistics

5.2.1.1 Oral test

Tables 5-10 and 5-11 below summarize the number of participants, minimum and maximum, mean scores, and standard deviations for CHL learner groups and non-CHL learner groups during the pre-test. Figure 5-4 graphically presents the mean scores of the two learners groups respectively. For CHL learners, the means scores of learners in
the recast group ($M=58.87$) and those in the meta-linguistic group ($M=57.22$) was very close. They scored higher than learners in the control group ($M=41.66$). On the other hand, for non-CHL learners, learners in the recast group ($M=19.63$) scored slightly higher than those in the meta-linguistic group ($M=15.8$). The mean scores of learners in the control group ($M=16.44$) were between the scores of learners in the two experimental groups.

Table 5-10. Summary of oral pre-test scores for CHL learner group

<table>
<thead>
<tr>
<th>Group</th>
<th>$N$</th>
<th>Minimum</th>
<th>Maximum</th>
<th>$M$</th>
<th>$SD$</th>
</tr>
</thead>
<tbody>
<tr>
<td>RHL</td>
<td>12</td>
<td>38</td>
<td>92</td>
<td>58.87</td>
<td>18.28</td>
</tr>
<tr>
<td>MHL</td>
<td>11</td>
<td>23</td>
<td>90</td>
<td>57.22</td>
<td>19.15</td>
</tr>
<tr>
<td>CHL</td>
<td>8</td>
<td>10</td>
<td>85</td>
<td>41.66</td>
<td>27.75</td>
</tr>
</tbody>
</table>

Table 5-11. Summary of oral pre-test scores for non-CHL learner group

<table>
<thead>
<tr>
<th>Group</th>
<th>$N$</th>
<th>Minimum</th>
<th>Maximum</th>
<th>$M$</th>
<th>$SD$</th>
</tr>
</thead>
<tbody>
<tr>
<td>NRHL</td>
<td>12</td>
<td>6</td>
<td>56</td>
<td>19.63</td>
<td>13.32</td>
</tr>
<tr>
<td>NMHL</td>
<td>11</td>
<td>4</td>
<td>48</td>
<td>15.80</td>
<td>13.13</td>
</tr>
<tr>
<td>NCHL</td>
<td>9</td>
<td>10</td>
<td>21</td>
<td>16.44</td>
<td>3.67</td>
</tr>
</tbody>
</table>

Figure 5-4. Mean scores of oral pre-test

Tables 5-12 and 5-13 below summarize the descriptive statistics by CHL learner groups and non-CHL learner groups during the post-test. The means scores of learners in all the groups are also graphically presented in figure 5-5. For CHL learners, those in
the recast group ($M=73.45$) scored slightly lower than those in the meta-linguistic group ($M=78.05$). Learners in both experimental groups scored much higher than learners in the control group ($M=50.54$). On the other hand, for non-CHL learners, learners in the recast group ($M=35.07$) also scored slightly lower than learners in the meta-linguistic group ($M=36.64$). Learners in the control group ($M=15.52$) scored much lower than learners in the two experimental groups.

### Table 5-12. Summary of oral post-test scores of CHL learner group

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>Minimum</th>
<th>Maximum</th>
<th>$M$</th>
<th>$SD$</th>
</tr>
</thead>
<tbody>
<tr>
<td>RHL</td>
<td>12</td>
<td>42</td>
<td>92</td>
<td>73.45</td>
<td>16.53</td>
</tr>
<tr>
<td>MHL</td>
<td>11</td>
<td>23</td>
<td>94</td>
<td>78.05</td>
<td>21.18</td>
</tr>
<tr>
<td>CHL</td>
<td>8</td>
<td>19</td>
<td>83</td>
<td>50.54</td>
<td>22.59</td>
</tr>
</tbody>
</table>

### Table 5-13. Summary of oral post-test scores of non-CHL learner group

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>Minimum</th>
<th>Maximum</th>
<th>$M$</th>
<th>$SD$</th>
</tr>
</thead>
<tbody>
<tr>
<td>RNHL</td>
<td>12</td>
<td>10</td>
<td>75</td>
<td>35.07</td>
<td>23.24</td>
</tr>
<tr>
<td>MNHL</td>
<td>12</td>
<td>6</td>
<td>40</td>
<td>36.64</td>
<td>11.52</td>
</tr>
<tr>
<td>CNHL</td>
<td>9</td>
<td>6</td>
<td>90</td>
<td>15.52</td>
<td>21.54</td>
</tr>
</tbody>
</table>

![Figure 5-5. Mean scores of oral post-test](image)

5.2.1.2 Written test

Tables 5-14 and 5-15 below summarize the descriptive statistics by CHL learner groups and non-CHL learner groups during the pre-test. For CHL learners, learners in
the recast group \((M=36.12)\) scored relatively higher than those in the meta-linguistic group \((M=30.50)\). The mean scores of learners in the control group \((M=29.60)\) were close to the mean scores of learners in the meta-linguistic group. On the other hand, for non-CHL learners, learners in recast group \((M=26.56)\) scored slightly higher than those in the meta-linguistic group \((M=21.54)\). Learners in the control group \((M=16.89)\) scored relatively lower than learners in the two experimental groups. Figure 5-6 graphically present the mean scores of all the groups.

Table 5-14. Summary of written pre-test scores for CHL learner group

<table>
<thead>
<tr>
<th>Group</th>
<th>(N)</th>
<th>Minimum</th>
<th>Maximum</th>
<th>(M)</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>RHL</td>
<td>12</td>
<td>7</td>
<td>76</td>
<td>36.12</td>
<td>21.66</td>
</tr>
<tr>
<td>MHL</td>
<td>11</td>
<td>0</td>
<td>62</td>
<td>30.50</td>
<td>20.08</td>
</tr>
<tr>
<td>CHL</td>
<td>8</td>
<td>2</td>
<td>59</td>
<td>29.60</td>
<td>21.33</td>
</tr>
</tbody>
</table>

Table 5-15. Summary of written pre-test scores for non-CHL learner group.

<table>
<thead>
<tr>
<th>Group</th>
<th>(N)</th>
<th>Minimum</th>
<th>Maximum</th>
<th>(M)</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>RNHL</td>
<td>12</td>
<td>4</td>
<td>61</td>
<td>26.56</td>
<td>15.88</td>
</tr>
<tr>
<td>MNHL</td>
<td>12</td>
<td>3</td>
<td>50</td>
<td>21.54</td>
<td>13.76</td>
</tr>
<tr>
<td>CNHL</td>
<td>9</td>
<td>1</td>
<td>30</td>
<td>16.89</td>
<td>9.67</td>
</tr>
</tbody>
</table>

Figure 5-6. Mean scores of written pre-test

Tables 5-16 and 5-17, below, summarize the descriptive statistics by CHL learner groups and non-CHL learner groups during the written post-test. For CHL learners,
learners in the recast group ($M=53.14$) scored relatively higher than those in the meta-linguistic group ($M=48.79$). Learners in the control group ($M=33.89$) scored much lower than those in both two experimental groups. On the other hand, for non-CHL learners, learners in the recast group ($M=36.76$) scored relatively lower than learners in the meta-linguistic group ($M=46.72$). Learners in the control group ($M=17.06$) scored much lower than learners in the two experimental groups. Figure 5-7 graphically present the mean scores of all the groups.

Table 5-16. Summary of written post-test scores for CHL learner group

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>Minimum</th>
<th>Maximum</th>
<th>$M$</th>
<th>$SD$</th>
</tr>
</thead>
<tbody>
<tr>
<td>RHL</td>
<td>12</td>
<td>16</td>
<td>88</td>
<td>53.14</td>
<td>18.84</td>
</tr>
<tr>
<td>MHL</td>
<td>11</td>
<td>7</td>
<td>72</td>
<td>48.79</td>
<td>19.02</td>
</tr>
<tr>
<td>CHL</td>
<td>8</td>
<td>4</td>
<td>67</td>
<td>33.89</td>
<td>24.76</td>
</tr>
</tbody>
</table>

Table 5-17. Summary of written post-test scores for non-CHL learner group

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>Minimum</th>
<th>Maximum</th>
<th>$M$</th>
<th>$SD$</th>
</tr>
</thead>
<tbody>
<tr>
<td>RNHL</td>
<td>12</td>
<td>13</td>
<td>70</td>
<td>36.76</td>
<td>16.76</td>
</tr>
<tr>
<td>MNHL</td>
<td>12</td>
<td>11</td>
<td>80</td>
<td>46.72</td>
<td>21.56</td>
</tr>
<tr>
<td>CNHL</td>
<td>9</td>
<td>2</td>
<td>29</td>
<td>17.06</td>
<td>8.89</td>
</tr>
</tbody>
</table>

Figure 5-7. Mean scores of written post-test
In summary, the descriptive statistics of the oral and written pre-tests showed that learners in the recast group scored slightly higher than learners in the meta-linguistic group. That indicated that learners in the recast group had a slight advantage over those in the meta-linguistic group before they received any feedback during the treatment. However, during the post-test, learners in the recast group were outperformed by those in the meta-linguistic group. The only exception occurred for CHL learners in the written test: learners in the recast group outperformed those in the meta-linguistic feedback group. These comparisons are summarized in table 5-18 below. Inferential statistics were employed to test whether any group differences existed in both pre- and post-tests. The results will be presented in 5.2.2.

<table>
<thead>
<tr>
<th>Test</th>
<th>Pre</th>
<th>Post</th>
<th>Pre</th>
<th>Post</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oral</td>
<td>RHL&gt;MHL&gt;CHL</td>
<td>MHL&gt;RHL&gt;CHL</td>
<td>RNHL&gt;CNHL&gt;MNHL</td>
<td>MNHL&gt;RNHL&gt;CNHL</td>
</tr>
<tr>
<td>Written</td>
<td>RHL&gt;MHL&gt;CHL</td>
<td>RHL&gt;MHL&gt;CHL</td>
<td>RNHL&gt;MNHL&gt;CNHL</td>
<td>MNHL&gt;RNHL&gt;CNHL</td>
</tr>
</tbody>
</table>

5.2.1.3 Learners' previous knowledge reflected in their verbal comments

Tables 5-19 and 5-20 summarize learners’ previous knowledge of the target classifiers based on their verbal reports collected in the stimulated recall. Figures 5-8 and 5-9 graphically present the results. The percentage of (re)-acquisition is approximately two times higher in the CHL learner group ($M=14.0\%$) than in the non-CHL learner group ($M=6.8\%$), indicating that CHL learners had knowledge of these classifiers before. CHL learners also present a higher percentage in the category of A2 ($M=48.8\%$) than non-CHL learners ($M=35.2\%$), suggesting that they knew a wider range of classifiers. A slightly higher percentage for the category of A1 in non-CHL groups ($M=31.3\%$) indicated that they were more likely to use the general classifier ge instead
of the targeted classifier. They also tended to skip classifiers \( (M=26.7\%) \) more often than the CHL learners \( (M=10.3\%) \), which is reflected in the category of O.

Table 5-19. CHL learners’ previous knowledge of target classifiers

<table>
<thead>
<tr>
<th>Group</th>
<th>R (%)</th>
<th>A1 (%)</th>
<th>A2 (%)</th>
<th>O (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>RHL(n=155)</td>
<td>9 (5.8)</td>
<td>63 (40.6)</td>
<td>70 (45.2)</td>
<td>13 (8.4)</td>
</tr>
<tr>
<td>MHL(n=166)</td>
<td>36 (21.7)</td>
<td>24 (14.5)</td>
<td>86 (51.8)</td>
<td>20 (12.0)</td>
</tr>
<tr>
<td>Total(n=320)</td>
<td>45 (14.0)</td>
<td>87 (27.2)</td>
<td>156 (48.8)</td>
<td>33 (10.3)</td>
</tr>
</tbody>
</table>

Note: R=(re)acquisition, A1=substitution of ge, A2=wide range of (non-target/incorrect) classifiers, O=skip a classifier

Table 5-20. Non-CHL learners’ previous knowledge of target classifiers

<table>
<thead>
<tr>
<th>Group</th>
<th>R (%)</th>
<th>A1 (%)</th>
<th>A2 (%)</th>
<th>O (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>RNHL(n=217)</td>
<td>16 (7.4)</td>
<td>96 (44.2)</td>
<td>69 (31.8)</td>
<td>36 (16.6)</td>
</tr>
<tr>
<td>MNHL(n=266)</td>
<td>17 (6.4)</td>
<td>55 (20.7)</td>
<td>101 (38.0)</td>
<td>93 (35.00)</td>
</tr>
<tr>
<td>Total(n=483)</td>
<td>33 (6.8)</td>
<td>151 (31.3)</td>
<td>170 (35.2)</td>
<td>129 (26.7)</td>
</tr>
</tbody>
</table>

Figure 5-8. CHL learners’ previous knowledge of target classifiers

Figure 5-9. Non-CHL learners’ previous knowledge of target classifiers
5.2.2 Inferential Statistics

Learners’ pre-test scores were submitted to the one-way analysis of variance (ANOVA) measure to test for group equivalence. No significant group differences were found in pre-test scores across all experimental and control groups, indicating group equivalence at the onset of the study. The results are summarized in tables 5-21 and 5-22 below.

Table 5-21. One-way ANOVA for the oral pre-test

<table>
<thead>
<tr>
<th>Group</th>
<th>Df</th>
<th>MS</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>CHL</td>
<td>2</td>
<td>807.54</td>
<td>1.78</td>
<td>.19</td>
</tr>
<tr>
<td>Non-CHL</td>
<td>2</td>
<td>49.38</td>
<td>.38</td>
<td>.69</td>
</tr>
</tbody>
</table>

P<.05

*Note. CHL = includes CHL learners in all the experimental and control groups Non-CHL = includes non-CHL learners in all the experimental and control groups

Table 5-22. One-way ANOVA for the written pre-test

<table>
<thead>
<tr>
<th>Group</th>
<th>Df</th>
<th>MS</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>CHL</td>
<td>2</td>
<td>49.38</td>
<td>.38</td>
<td>.69</td>
</tr>
<tr>
<td>Non-CHL</td>
<td>2</td>
<td>134.07</td>
<td>.30</td>
<td>.74</td>
</tr>
</tbody>
</table>

P<.05

In order to answer the research question whether explicit feedback works more effectively than implicit feedback for CHL and non-CHL learners’ language development, a one-way ANOVA measure was employed to examine group differences in the post-test.

The independent variable was the feedback type received by learners during the treatment. The dependent variable was their scores during the pre- and post-tests. The scores of each learner group were separately submitted to the one-way ANOVA. The alpha level was set at .05. A post hoc LSD test was conducted to locate the source of statistical significance. Learners’ scores on the post-test were submitted to the one-way ANOVA measure to examine whether there was a significant difference between
groups. Significant between-group differences were found in both the oral and written post-test for the non-CHL learner group, as well as in the oral post-test for the CHL learner group. No significant between-group differences were found for the CHL learner group in the written post-test. The results are summarized in tables 5-23 and 5-24 below.

Table 5-23. One-way ANOVA for the oral post-test

<table>
<thead>
<tr>
<th>Group</th>
<th>Df</th>
<th>MS</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>CHL</td>
<td>2</td>
<td>1931.85</td>
<td>4.89</td>
<td>.015*</td>
</tr>
<tr>
<td>Non-CHL</td>
<td>2</td>
<td>1360.35</td>
<td>3.37</td>
<td>.048*</td>
</tr>
</tbody>
</table>

Table 5-24. One-way ANOVA for the written post-test

<table>
<thead>
<tr>
<th>Group</th>
<th>Df</th>
<th>MS</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>CHL</td>
<td>2</td>
<td>2221.94</td>
<td>7.70</td>
<td>.13</td>
</tr>
<tr>
<td>Non-CHL</td>
<td>2</td>
<td>929.58</td>
<td>2.20</td>
<td>.002*</td>
</tr>
</tbody>
</table>

Tables 5-25 and 5-26 summarize the results of the LSD post-hoc test for the oral post-test. CHL learners in the two experimental groups significantly outperformed those in the control group. However, statistical significance was not found between the two experimental groups (p=.584). Similar results were found for non-CHL learners.

Table 5-25. LSD post hoc test for CHL learners on the oral test

<table>
<thead>
<tr>
<th>Pairs</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>RHL-MHL</td>
<td>.584</td>
</tr>
<tr>
<td>RHL-CHL</td>
<td>.018*</td>
</tr>
<tr>
<td>MHL-CHL</td>
<td>.006*</td>
</tr>
</tbody>
</table>

Table 5-26. LSD post hoc test for non-CHL learners on the oral test

<table>
<thead>
<tr>
<th>Pairs</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>RNHL-MNHL</td>
<td>.849</td>
</tr>
<tr>
<td>RNHL-CNHL</td>
<td>.035*</td>
</tr>
<tr>
<td>MNHL-CNHL</td>
<td>.024*</td>
</tr>
</tbody>
</table>

P<.05
Tables 5-27 and 5-28 summarize the results of the LSD post-hoc test for the written post-test. Table 5-27 showed that CHL learners in the control group were significantly outperformed by those in the recast group (p=.049), but not by those in the meta-linguistic group. On the other hand, non-CHL learners in the two experimental groups did significantly better than those in the control group. However, no significant results were found between the two experimental groups for either CHL or non-CHL learners.

Table 5-27. LSD post hoc test for CHL learner on the written test

<table>
<thead>
<tr>
<th>Pairs</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>RHL-MHL</td>
<td>.616</td>
</tr>
<tr>
<td>RHL-CHL</td>
<td>.049*</td>
</tr>
<tr>
<td>MHL-CHL</td>
<td>.130</td>
</tr>
</tbody>
</table>

P<.05

Table 5-28. LSD post hoc test for non-CHL learner on the written test

<table>
<thead>
<tr>
<th>Pairs</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>RNHL-MNHL</td>
<td>.183</td>
</tr>
<tr>
<td>RNHL-CNHL</td>
<td>.013*</td>
</tr>
<tr>
<td>MNHL-CNHL</td>
<td>.001*</td>
</tr>
</tbody>
</table>

P<.05

In summary, the inferential statistics revealed significant between-group differences across the experimental and control groups regardless of their CHL background. Overall, learners in the experimental groups did significantly better than learners in the control groups, but no significant between-group differences were found between the recast group and the meta-linguistic feedback group. However, in the written post-test, CHL learners in the recast group significantly outperformed the control group, whereas those in the meta-linguistic feedback group did not do significantly better than the control group.
5.3 Research Question 3

Research question 3 asked whether feedback type or language background affects the increase in CHL and non-CHL learners’ performance on a) an oral imitation test, b) an untimed written cloze test.

5.3.1 Descriptive Statistics

5.3.1.1 Oral imitation test

Tables 5-29 and 5-30 present the number of participants, minimum, and maximum scores, as well as the mean scores of the increase made by learners in all the experimental groups. The increase was calculated by the score of the post-test minus that of the pre-test. Figure 5-10 shows a graphical representation of the increase by learners in the experimental groups. All the learners made an increase from pre-test to post-test. In the recast group, CHL learners and non-CHL learners made exactly the same increase ($M=15.33$). In the meta-linguistic groupop, there was also little difference between the two learner groups. In terms of feedback condition, learners in the meta-linguistic feedback group made a greater increase than those in the recast group by approximately five points.

Table 5-29. Summary of increase on the oral post-test for recast group

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>Minimum</th>
<th>Maximum</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>RHL</td>
<td>12</td>
<td>-4</td>
<td>32</td>
<td>15.33</td>
<td>12.38</td>
</tr>
<tr>
<td>RNHL</td>
<td>12</td>
<td>-4</td>
<td>35</td>
<td>15.33</td>
<td>12.01</td>
</tr>
</tbody>
</table>

Table 5-30. Summary of increase on the oral post-test for meta-linguistic group

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>Minimum</th>
<th>Maximum</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>MHL</td>
<td>11</td>
<td>-2</td>
<td>42</td>
<td>20.91</td>
<td>15.31</td>
</tr>
<tr>
<td>MNHL</td>
<td>12</td>
<td>2</td>
<td>44</td>
<td>21.00</td>
<td>13.14</td>
</tr>
</tbody>
</table>
Figure 5-10. Increase on the oral post-test across the experimental groups

Table 5-31 presents the descriptive statistics of learners in the control groups. The results show that the increase made by CHL learners in the control group was 9 points, which indicated that these learners showed an improvement of the target form by simply taking the tests. In contrast, the increase of non-CHL learners was minus 1.11 point, which indicated that these learners did not make any progress by taking the tests.

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>Minimum</th>
<th>Maximum</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>CHL</td>
<td>8</td>
<td>-2</td>
<td>21</td>
<td>9.00</td>
<td>8.00</td>
</tr>
<tr>
<td>CNHL</td>
<td>9</td>
<td>-13</td>
<td>19</td>
<td>-1.11</td>
<td>10.24</td>
</tr>
</tbody>
</table>

### 5.3.1.2 Written cloze test

Tables 5-32 and 5-33 present the descriptive statistics of data on the post-test of learners in the experimental groups. Figure 5-11 shows a graphical representation of increases made by learners in all the experimental groups. The results show that in the recast group the CHL learners made greater increases (\(M = 15.33\)) than the non-CHL learners (\(M = 9.17\)). In contrast, in the meta-linguistic group the non-CHL learners (\(M = 21.92\)) outperformed the CHL learners (\(M = 16.45\)).
Table 5-32. Summary of increases on the written post-test for recast group

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>Minimum</th>
<th>Maximum</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>RHL</td>
<td>12</td>
<td>3</td>
<td>33</td>
<td>15.33</td>
<td>9.30</td>
</tr>
<tr>
<td>RNHL</td>
<td>12</td>
<td>-3</td>
<td>18</td>
<td>9.17</td>
<td>5.62</td>
</tr>
</tbody>
</table>

Table 5-33. Summary of increases on the written post-test scores for meta-linguistic group

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>Minimum</th>
<th>Maximum</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>MHL</td>
<td>11</td>
<td>4</td>
<td>34</td>
<td>16.45</td>
<td>9.92</td>
</tr>
<tr>
<td>MNHL</td>
<td>12</td>
<td>2</td>
<td>36</td>
<td>21.92</td>
<td>11.18</td>
</tr>
</tbody>
</table>

Table 5-34 presents the descriptive statistics of learners in the control groups. The results show that the increase made by CHL learners in the control group was 4.13 points, which indicated that these learners showed a slight improvement of the target form by simply taking the tests. In contrast, the increase of non-CHL learners was only .56 point, which indicated that these learners did not benefit from taking the tests.

Figure 5-11. Increases on the written post-test across the experimental groups

Table 5-34. Summary of increases on the written post-test for control group

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>Minimum</th>
<th>Maximum</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>CHL</td>
<td>8</td>
<td>-7</td>
<td>12</td>
<td>4.13</td>
<td>6.47</td>
</tr>
<tr>
<td>NCHL</td>
<td>9</td>
<td>-6</td>
<td>7</td>
<td>.56</td>
<td>4.16</td>
</tr>
</tbody>
</table>

In summary, the descriptive statistics results show that learners in the meta-linguistic feedback group generally made higher increase in both oral and written tests regardless of their language background. Interestingly, whereas CHL learners in the
control group made improvements in their increase in both oral and written tests, non-CHL learners made little or even regressed improvement.

5.3.2 Inferential Statistics

In order to answer the question of whether different feedback types led to an increase in different types of knowledge, a 2 X 2 full-factorial analysis of variance (ANOVA) was conducted to determine the effects of feedback type and learners’ language background on their increase. The use of a factorial ANOVA design in the analysis of between-group differences was motivated by the following fact: in the current study, there are two independent variables that split the participants into separate groups: the feedback type and the CHL background of the learners. Therefore we can also call it a two-way between-groups design.

I will first report the results for the oral test. The main effect of feedback type, ignoring the CHL background variable, is not statistical ($F_{1,43} = 2.123, p = .152$, partial eta-squared= .047 , power= .297). It indicates that learners who received meta-linguistic feedback did not increase a significantly higher score on the oral post-test than those who received recasts during the treatment. Similarly, it was found that learners’ CHL background did not produce a statistically significant effect for the main effect of CHL background ($F_{1,43} = .000, p = .991$, partial eta-squared= .000 , power= .050), indicating that overall, learners made similar increase regardless of their language background. The interaction between feedback type and CHL background was not found to be statistically significant either.

Next I will report the results of the written test. The main effect of feedback type, ignoring CHL background variable, is statistically significant ($F_{1,43} = 6.888, p = .012$, partial eta-squared= .138 , power= .728). It indicates that learners who received meta-
linguistic feedback made a greater increase in the written post-test than those who received recasts during the treatment. The effect size showed that this factor accounted for 72.8% of the variance in the data, which is a relatively large effect. In contrast, the main effect of the CHL background variable, is not statistically significant ($F_{1,43} = .032, p = .859$, partial eta-squared= .01 , power= .054). However, the interaction between feedback type and CHL background was found to be statistically significant ($F_{1,43} = 4.663, p = .036$, partial eta-squared= .098 , power= .560). The effect size showed that the interaction factor accounted for 56.0% of the variance in the data, which is a relatively small effect. Thus we really want to know whether the combinations of feedback and learners’ language background are statistically different from each other. A post hoc test was employed with an adjusted $p$-value of .012. The result showed that for the CHL learners, the feedback type was not very important ($p= .772$). In contrast, for the non-CHL learners, the feedback type was very important ($p= .002$): those in the meta-linguistic feedback group performed significantly better than those in the recast group.

---

$^{1}$ .012 is the cut-off value of False Detection Rate (FDR) calculated by R statistical program.
CHAPTER 6
DISCUSSION AND CONCLUSION

In this chapter, I will discuss the findings of the study with regards to the research questions, followed by a consideration of the theoretical and pedagogical implications. Then I will address some limitations of the study, and suggest areas for future research. Finally, I will conclude the chapter by addressing how the current study extends our knowledge about the effects of implicit and explicit feedback.

6.1 Research Question 1

Research question 1 asked whether CHL learners perceive both implicit and explicit feedback on Chinese classifiers more accurately than non-CHL learners.

In this study of learners’ perceptions about target classifiers in conversational interaction, learners with CHL background did not show any statistically significant advantage over those without CHL background in the same feedback condition. The only significant difference between the two learner groups was seen between the CHL learners in the meta-linguistic feedback group and the non-CHL learners in the recast group. On the other hand, learners in the meta-linguistic feedback group all significantly outperformed those in the recast group, regardless of their language background.

Comparing the findings of current study with those of previous research (e.g., Lyster & Ranta, 1997; Lyster 1998a, 1998b; Panova & Lyster, 2002), it is noticeable that the percentage of learners’ accurate perception of feedback was higher, particularly in the case of recasts (57%~88%). The discrepancy of results between the current study and previous studies could be due to the following factors: the learning context; the learners’ orientation, and their learning experience.
First of all, previous research found that learning context could play an important role in learners’ perception of recasts. For instance, recasts were more likely to be noticed in a FL classroom rather than in an immersion classroom. This is because in a FL classroom, linguistic forms are usually emphasized more often; whereas in an immersion classroom, instructors focus more on communication and meaning (R. Ellis & Sheen, 2006; Lyster & Mori, 2006; Nicholas et al., 2001; Sheen, 2004). The current study was conducted in a laboratory setting, in which recasts were delivered intensively on the target forms in a dyad between a learner and the NS interlocutor. As a result, recasts in the current study might have become less implicit. Therefore, they were more easily noticed by the learners, regardless of their language background. The stimulated recall data in example 6-1 illustrate that, despite the distractors used in the study and the corrections provided for all errors, the repeated appearances of target classifiers raised learners’ awareness of recasts for these structures.

Example 6-1  A non-CHL learners’ verbal comments

NNS: *我买了四个笔。
    Wo mai le si ge bi
    1sg buy PFV four CL pen
    I bought four pens.

NS: 你买了四枝笔？
    Ni mai le si zhi bi
    2sg buy PFV four CL pen
    You bought four pens?

NNS: 四枝笔。
    Si zhi bi
    Four CL pen
    Four pens.

Recall data
Cause when you asked, and I said, and you said it after. I tried to trouble-listen for the measure words. So I would know. Cause I don't want to say a wrong measure word. People may be like: what you are talking about?
In Example 6-1, the learner reported that he noticed he was frequently corrected by the NS interlocutor by her reformulation of his original words. In addition, he focused his attention to trouble-listen particularly for the target classifiers.

In summary, the laboratory setting employed in the current study might have made recasts less implicit than was expected. As a result, they were more easily noticed by learners, regardless of their language background. Although the CHL learners (71.43%) still showed an advantage over the non-CHL learners (57.26%) in their perceptions, the difference was not great enough to be statistically significant.

Secondly, learners’ orientation may have also affected their perceptions of feedback. Generally, HL learners share a similar pattern in their perceptions of feedback due to their exposure to the HL in the early years. Even though the length and the extent of the exposure are limited, the effect of the experience still lasts for a very long time (Au et al., 2002). When engaging in a conversational interaction with a NS interlocutor, learners with HL background are more likely to focus on lexical and semantic issues compared with those without HL background (Gass & Lewis, 2007). In addition, they tend to treat the interaction as a real-life communication that they have been experiencing with family and community members. In contrast, learners without HL experience are more likely to consider the interaction as a language exercise that they always experience in a classroom (Kim, 2008). Therefore, HL learners could be considered more meaning-oriented, whereas non-HL learners are more form-oriented. Due to this different orientation, HL learners perceived feedback on semantic and lexical items more accurately than non-HL learners in Gass and Lewis’s study (2007). This finding was supported by the descriptive statistics of current study (see tables 5-5 and
5-6), given that classifiers were also semantic and lexical in nature. However, Gass and Lewis neither examined the relationship between feedback type and learners’ perception, nor employed inferential statistics to assess the significance of their results. These two gaps were filled by the current study.

In the current study, the CHL learners in the meta-linguistic feedback condition were significantly more accurate in their perceptions of feedback than the non-CHL learners in the recast condition \((p=.000)\). In addition, within the CHL learner group, those in the meta-linguistic feedback group also performed significantly better than those in the recast group \((p=.003)\). If learners’ reports about their perceptions of feedback can be equated with attention, then this result could be interpreted as providing evidence that meta-linguistic feedback is superior to recasts in helping CHL learners temporarily shift their attention from meaning-oriented communication to the target linguistic forms. A CHL learner’s verbal comment in Example 6-2 provides support for this interpretation.

Example 6-2 A CHL learners’ verbal comments

NNS: *一条花？
Yi tiao hua
One CL flower
One flower?

NS: 这么一个整个的花由茎连起来，
Zheme yi ge zhengge de hua you jing lian qilai
This kind one CL whole NOM flower by stem connect CP
This kind of one whole flower (that was) connected by a stem
叫什么？
Jiao shenme
Call what
What (do we) call (it)?

NNS: *条。
Tiao
CL
Tiao

Recall data

1不用的这个字我都忘了。平常我不说花呀、朵呀。平常跟父母在一起说中文的时候，就说出来了。我不想我是不是叫“朵”，叫“枝”还是叫“面”。你让我在这想他们叫什么。很多我就忘了。
(I forgot the word that I did not use much. In daily life I don’t talk about flowers much. When I speak Chinese with my parents, I can say them (such as flowers). I do not think whether I (should) call it duo, zhi, or mian. You made me think what they (should be) called. I forgot many of (the measure words).

In Example 6-2, the CHL learner’s verbal comments implied that he usually does not pay much attention to his choice of words when speaking with his parents in Chinese. However, the meta-linguistic feedback given by the interlocutor forced him to attend to the appropriate linguistic form, which he had acquired in the past but subsequently forgot.

In contrast, recasts could sometimes be interpreted as an alternative way for saying the same thing with the purpose of keeping the communicative channel open. Example 6-3 illustrated this.

Example 6-3 CHL learner XSB’s verbal comment

NNS: *房间放着一张镜子。
Fangjian fang zhe yi zhang jingzi
Room put CP One CL mirror
There is a mirror in the room.

NS: 那面镜子高不高？
Na mian jingzi gao bu gao
That CL mirror tall not tall
Is that mirror tall or not?

NNS: *不高。
Bu gao
Not tall

Recall data

2 That one is OK. I forgot for a second the word mirror for Chinese.
In Example 6-3, the CHL learner did not perceive the provided recast as a corrective feedback, which was reflected in his response toward it: instead of producing an immediate uptake to modify the non-targetlike classifier, he simply responded to continue the interaction. A post hoc analysis was conducted to investigate this learner’s immediate uptake pattern after feedback during the interaction. Uptake here refers to “the learners’ modification of their original utterance following the NS’s provision of feedback through recasts or meta-linguistic feedback” (Mackey et al., p. 492). As discussed in Chapter 3, uptake may not be a reliable measure of noticing of the feedback, particularly in the case of recasts. However, it may be related to learners’ perceptions about feedback at the time of the feedback. The result showed that learner XSB tended to provide uptake after receiving feedback. In addition, for those recasts for which he did not provide uptake, none was accurately perceived. These results are shown in tables 6-1.

<table>
<thead>
<tr>
<th>+Uptake (n=18)</th>
<th>-Uptake (n=4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perceived</td>
<td>Not perceived</td>
</tr>
<tr>
<td>10</td>
<td>8</td>
</tr>
<tr>
<td>(55.56%)</td>
<td>(44.44%)</td>
</tr>
</tbody>
</table>

In addition, a second post hoc analysis showed that learner XSB used the same problematic classifier *zhang* in both his oral and written pre- and post-tests. All this evidence suggested that this learner assumed *zhang* was the correct classifier for *mirror*. The target classifier *mian* provided by the interlocutor at the moment of the interaction was apparently assumed to be just an alternative one by this learner, and never assimilated into his interlanguage system.
In summary, meta-linguistic feedback overtly indicates the error with meta-linguistic information, thus it helps CHL learners shift their attention from meaning-oriented communication to the target form. Compared with meta-linguistic feedback, recasts are less likely to draw learners’ attention to the target form, particularly in task-based or content-based lessons, in which meaning and communication, rather than linguistic forms, are in focus (Lyster, 1998a, 1998b; Lyster & Ranta, 1997).

Thirdly, learner’s previous learning experience might also have had impact on their perceptions of feedback, particularly recasts. When Chinese L1 children acquire classifiers at an early age, they tend to treat the classifier and noun construction as an inseparable linguistic unit and learn them by simple memorization (Fang, 1985). This finding was further supported by learners’ verbal comments collected during the stimulated recall in the current study. Examples 6-4 and 6-5 below are two of the representative verbal reports.

Example 6-4  At home
NNS: *也买了三只皮带。
Ye mai le san zhi pidai
Also buy PFV three CL belt
(I) also bought three belts.
NS: 那三条皮带是什么颜色的？
Na san tiao pidai shi shenme yanse de
That three CL belt is what color NOM
What color are those three belts?
NNS: 三条皮带，
San tiao pidai
Three CL belt
Three belts,
红色黑色的，
Hongse heise de
Red black NOM
Red and black,
还有一个灰色的。
Hai you yi ge huise de
Also have one CL gray NOM
(There is) also a gray one.

Recall data
皮带，皮带的量词。一条皮带。我也不知道皮带的量词是什么，我没有穿过皮带。很多词如果我是常用的东西，像一枝笔，一张纸，我会知道这个量词是什么。可是皮带我没有穿过什么皮带，所以我不知道一条皮带。
(Belts, the classifier for belts. One belt. I didn’t know the classifier for belts either. I have never worn belts before. For things I often use, such as one pen, a piece of paper, I would know the classifiers for them. But I have not really worn belts, so I didn’t know (how to say) one belt).

Example 6-5 At community Chinese language school
NNS: *一 个 蛇。
Yi ge she
One CL snake
NS: 那 条 蛇 是 什么 颜色 的?
Na tiao she shi shenme yanse de?
That CL snake is what color NOM
What color is that snake?
NNS: 那条蛇。白色，黄色的。
Na tiao She. Baise, huangse de
That CL snake white yellow
That snake (is) white, yellow.

Recall data
小时候我妈送我去中文学校，很多家长会教他们的孩子怎么说中文，我记得我上了一堂课，一门课，我上了一堂课，老师给了我们一个poem, 我早忘了是说什么，可是我记得是讲动物园。所以我记得一头牛，一匹马，一只猫。
(My mom sent me to a Chinese school when I was young, where many parents taught their children how to speak Chinese. I remembered there was this lesson, in which the teacher taught us a poem. I forgot what the poem was about. But I remember it was about a zoo. So I remember how to say a cow ‘yi tou niu’, a horse ‘yi pi ma’, and a cat ‘yi zhi mao’).

The verbal comments in examples 6-4 and 6-5 were reported by the same CHL learner: Example 6-4 was about his learning experience at home, whereas Example 6-5 was drawn from his learning experience at a Chinese community school where he was
sent by his parents when he was young. Both examples indicated that he acquired classifiers and their associated nouns as a chunk. In other words, his learning experience of classifiers was overwhelmingly input-based. As a result, his knowledge was unbalanced. For instance, he knew the target classifier for horse, which was considered relatively difficult by the non-CHL learners in the current study. However, he did not know the less-difficult classifier for belt, as he did not wear belts often in daily life, and therefore did not have the opportunity to acquire the classifier for it.

The descriptive statistics results in tables 5-19 and 5-20 provided further evidence on this issue: the results showed that the CHL learners were more likely to use a non-target classifier that was supposed to associate with other nouns than the non-CHL learners (48.8% vs. 35.2%), which suggested that the CHL learners had heard a wider range of classifiers than the non-CHL learners. During the interaction, they mainly relied on these personal experiences to figure out whether a classifier was correct or not. Sometimes their experience was not accurate. Example 6-6 illustrates this.

Example 6-6
NNS: *还有两张镜子。

Haiyou liang zhang jingzi
Also two CL mirror
那里是 also two mirrors.

NS: “张”我们一般是说平面的东西，比如桌子。

Zhang women yiban shi shuo pingmian de dongxi
CL 1pl usually is say flat surface NOM thing
张, we usually (use it to) call something with a flat surface
但是说镜子我们不说“张”。
Danshi shuo jingzi women bu shuo zhang
But say mirror 1pl no say CL
zheng, we usually use it to describe an object with a flat surface,
But we don't use it for a mirror.

NNS: 镜子？

Jingzi
Mirror.

Recall data

“一张镜子”，我肯定听过，很熟。但是所以我想的时候，听不出来是错的。

肯定是小时候听过的。

(yizhang jingzi. I definitely have heard it, it sounded very familiar. Therefore when I was thinking (during the interaction), I could not see anything wrong there.)

In Example 6-6, the learner assumed zhang was the correct classifier because she had heard it in the past, even though the interlocutor clearly pointed out that her assumption was incorrect. A post hoc analysis showed that this learner still used zhang on her post-test.

In contrast, the acquisition of classifiers by non-CHL learners comes mainly from formal classroom instruction. Their knowledge of classifiers was found to relate positively to their proficiency levels: the higher their proficiency level in Chinese, the greater their knowledge of classifiers (Liang, 2008). Unlike CHL learners, they acquire classifiers through explicit learning rather than simply memorizing classifiers and their associated nouns as a unit. Example 6.7 illustrates a typical thinking process by a non-CHL learner toward the provided recast, which was reported during the stimulated recall.

Example 6-7 A non-CHL learners’ comment

NNS: *我也买一张,
Wo ye mai yi zhang
1sg also buy one CL
I also buy one,
一张大镜子。
Yi zhang da jingzi
One CL big mirror
One big mirror.

NS: 一面很大的镜子吗？
Yi mian hen da de jingzi ma
One CL very big NOM mirror QP
One big mirror?
NNS: 对。
    Dui.
    Yes.

Recall data
At first I thought the measure word would be *zhang* for *jingzi*, cause it is flat surface. But it wasn't, I was kind of surprised for a second, I heard you corrected me.

In Example 6-7, the learners’ attention was caught by the recast because he felt surprised about the target classifier provided by the interlocutor. Unlike CHL learners, he relied on the meta-linguistic knowledge that he had acquired through classroom learning rather than memorization of the classifier and an associated noun unit.

In summary, the previous learning experience helped both CHL learners and non-CHL learners to accurately perceive the recasts provided during the interaction. However, it worked in a different manner.

Long argued that feedback is beneficial for learners’ L2 development because it “connects input, internal learner capacities, particularly selective attention, and output in productive ways” (1996, 451-452). In order for feedback to play a facilitating role in promoting learning during interaction, learners need to focus their selective attention on linguistic forms and notice the gap between their problematic interlanguage and the target-like forms (Schmidt & Frota, 1996). If learners’ verbal reports of their perceptions can be quated with attention, it is interesting to know to which particular aspect of language learners attended when they failed to perceive feedback accurately. For the current study, it is particularly interesting to investigate this issue by comparing the two learner groups.
Regardless of the language background, when learners failed to make classifier-related comments, 85% of their comments were either unrelated to the feedback episode at all or related to lexis (see tables 5-7 and 5-8).

The descriptive statistics showed that the number of non classifier-related comments given by the CHL learners in the recast group (n=62) was approximately three times more than by CHL learners in the meta-linguistic group (n=20). A similar pattern was also seen with the non-CHL learners (recast: 162; meta-linguistic: 58). The gap between the two feedback groups indicated that learners who received recasts were undoubtedly less aware of the classifier-related feedback. In addition, the non-CHL learners were more likely to make no content recall comments compared with the CHL learners. Learners’ no content comments also showed that the CHL learners were more focused on telling a story, whereas the attention of the non-CHL learners was more easily captured by things that seemed unusual to them in the interaction.

Examples 6-8 and 6-9 illustrate this.

Example 6-8 A CHL learners’ no content comment

NNS: *我买一个帽子。
    Wo mai yi ge maozi
    1sg buy one CL hat
    I bought one hat.

NS: 那顶帽子是什么颜色的？
    Na ding maozi shi shenme yanse de
    That CL hat is what color NOM
    What color is that hat?

NNS: 白色的。
    Baise de
    White NOM.

Recall data
Yeah, I was just describing the picture.

Example 6-9 A non-CHL learners’ no content comment
Recall data
Like when you asked me can I see the dragon, I thought it was funny cause dragon is not real.

Although participants in the current study have already acquired or encountered all the stimuli nouns, as well as the majority of the vocabulary that emerged during the interaction (e.g., colors) through classroom instruction, many of them were still struggling due to a lack of adequate lexicon to describe the pictures. Although HL learners generally have a much wider range of vocabulary than non-HL learners, they may still have difficulties when describing things that are less frequently encountered in daily life.

Example 6-10 A CHL learners’ lexical comment

NNS: *我买了三个腰带。
Wo mai le san ge yaodai
1sg buy PFV three CL belt
I bought three belts,

NS: 那三条腰带是什么颜色？
Na san tiao yaodai shi shenme yanse
That three CL belt is what color
What color are (those) three belts?

NNS: 红色的、黑色的、还有…
Hongse de, heise de, haiyou
Red NOM black NOM also have
Red, black, there is also…

Recall data
I didn't know how to describe the color.

In Example 6-10, the CHL learner did not know how to describe the color of the belt, which was a light brown in the picture. In this particular case, the word for light brown was more important for maintaining the flow of communication compared to the classifier for belt. Not surprisingly, the learner’s verbal comments indicated that his attention was focused on how to describe the color rather than the classifier.

Compared with the CHL learners, the non-CHL learners have more difficulties in their vocabulary, which is illustrated in Example 6-11.

Example 6-11 A non-CHL learners’ lexical comment
NNS: *两个。
Liang ge
Two CL
Two.

NS: 这两只小鸡在做什么？
Zhe liang zhi xiaoji zai zuo shenme?
This two CL little chicken do what
What are the two little chickens doing?

NNS: 在，做，不知道。
Zai, zuo, bu zhidao
Pre do no know
(They are) doing, I don't know.

Recall data
I didn’t know the word for standing at all. And I didn’t know the word for chicken. I know I heard it before in China, but I didn’t put it together.

In Example 6-11, the non-CHL learner did not know the two most fundamental vocabulary items that were needed to carry on the conversation. Therefore her full attention was focused on the lexical issue rather than the classifier.

Besides the categories of no content and lexis, the two learner groups showed some differences in the distribution of categories of phonology, semantics, and the morphosyntax. The CHL learners were least likely to make a comment on phonology.
(only one case), which indicated their confidence about their pronunciation. In contrast, the non-CHL learners made most of their comments about phonology, particularly regarding the tone. Examples 6-12 and 6-13 illustrated these results.

Example 6-12 A CHL learner’ phonological comment
NNS: *有六，六朵花。
You liu duo hua
Have six CL flower
There are six flowers.
NS: 你的花带不带枝子？
Ni de hua dai bu dai zhizi?
You NOM flower have no have stem
Does your flower have a stem?
NNS: 有。
You
Have
(it) has.
NS: 那你有六枝花。
Na ni you liu zhi hua
Then 2sg have six CL flower
Then you have six flowers.

Recall data
I couldn't pronounce duo.

Example 6-13 A non-CHL learner’s phonological comment
NNS: *五个花。
Wu ge hua
Five CL flower
(There are) five flowers.
NS: 一二三四五，有六朵花。
Yi er san si wu liu, you liu duo hua
One two three four five six You six CL flower
(Count) one, two, three, four, five, six, there are six flowers.
NNS: Sorry。

Recall data
I got confused with huar (tone 1) and huar(tone 4).

On the other hand, the CHL learners in the recast group made most of their comments about morphosyntac issues. Interestingly, like the non-
CHL learners, they were also distracted by English word order. Examples 6-14 and 6-15 show two representative cases.

Example 6-14 A CHL learner’s morphosyntactic comment

NNS: *我买了一图画。
Wu ge hua
Five CL flower
(There are) five flowers.

NS: 你买了一张画。
Yi er san si wu liu, you liu duo hua
One two three four five six You six CL flower
(Count) one, two, three, four, five, six, there are six flowers.

NNS: 一张画。

Recall data
我在想应该放什么在画面的前面。因为在中文应该放像 adjective 一样的东西 to describe 东西的形状。英文就是放在后边。对我来说都是很难。
(I was thinking what word I should put in front of the word picture. As something like adjective should be put to describe the shape of an object in Chinese. It should be put after the word in English. It is both very difficult for me).

Example 6-15 A CHL learner’s morphosyntactic comment

NNS: *一个镜子。
Yi ge jingzi
One CL mirror
One mirror.

NS: 你那面镜子花了多少钱？
Ni na mian jingzi hua le duoshao qian
2sg that CL mirror spend PFV how much money
How much money did you spend for that mirror?

NNS: 旁边的钢琴。
Pangbian de gangqin
Next NOM piano
The piano next (to the mirror).

Recall data
I always forget the pattern you state the object that is something next to you first, like English is next to that table, but in Chinese, you say table next to is something. I always have to stop and think.
In summary, the CHL and the non-CHL learners tended to focus their attention on different language elements in the feedback when they did not perceive feedback as classifier-related. Regardless of their language background, they were more easily drawn to the areas in which they were less confident.

6.2 Research Question 2

Research question 2 asked whether explicit feedback works more effectively than implicit feedback for CHL and non-CHL learners’ language development. The results of the current study indicated that the answer to this research question is a ‘no’, which differs from findings of some previous research (e.g., Ammar & Spada, 2006; R. Ellis et al., 2006; Lyster, 2004). Therefore I will focus the discussion by asking why no significant differences emerged between feedback types among either the CHL learners or the non-CHL learners.

The one-way ANOVA on the oral post-test scores revealed identical results for the recast and the meta-linguistic feedback groups among both CHL and non-CHL learners. That is, regardless of the feedback type learners received during the treatment, they all significantly outperformed those in the control groups. Similar results were also found for non-CHL learners in the written test. Since group equivalence was not found at the onset, any increases by learners made in the post-test could be attributed to the effectiveness of the feedback they received during the treatment. These results suggested that both implicit and explicit feedback provided in communicative interactions were effective in promoting learners’ acquisition of the target linguistic forms. However, an unexpected finding in the current study was that no significant
between-group differences were found for CHL learners in the written test. The LSD post hoc test revealed that learners in the control group were significantly outperformed by learners in the recast group, but not by those in the meta-linguistic group ($p=.13$). I will discuss this issue at the end of this section.

Despite the significant differences existing between the experimental groups and the control group, no significant difference between the recast group and the meta-linguistic feedback group were found. These results echoed findings made in some previous studies (e.g., Carroll & Swain, 1993; DeKeyser, 2003; Kang, 2009; Loewen & Nabei, 2007; Nagata, 1993; Sanz, 2003; Sauro, 2009), but contrasted with results of other researchers (e.g., Ammar & Spada, 2006; Lyster, 2004; R. Ellis et al., 2006). Next I will explain why significant differences did not emerge between the two feedback conditions in the current study by comparing my results with some of the previous research.

The first explanation relates to the explicitness of the feedback provided during the interactions. As explained in Chapter 4, implicit feedback in the current study was operationalized in the form of recasts, which were always provided as implicitly as possible: they were always delivered in a complete sentence without any illocutionary cues (e.g., stress or higher pitch) to increase the salience of the reformulation. In contrast, explicit feedback was operationalized in the form of meta-linguistic feedback, which overtly indicated errors made by learners with specific meta-linguistic clues. Previous research found that the explicitness of feedback closely related to learners’ noticing of feedback. For instance, Rosa and Leow found that learners who received explicit information had higher levels of awareness than those who received more
implicit information on their target structures (2004). However, as discussed for research question 1, learners in the current study noticed more than half of the feedback, regardless of the type of feedback they received during the interactions. In other words, it seemed that recasts and meta-linguistic feedback worked equally effectively in drawing learners’ attention to the target linguistic forms. As discussed in section 6.1, the laboratory setting might have played an important role in increasing the likelihood of learners’ noticing, which supported the findings of previous studies (Han, 2002; Nicholas et al., 2001). Thus, recasts in the present study might have become more explicit compared with those in previous classroom studies that found significant differences between explicit and implicit feedback (e.g., Ammar & Spada, 2006; R. Lyster, 2004; R. Ellis et al., 2006).

Secondly, learners’ previous knowledge of the target classifiers could also contribute to the lack of difference between the two feedback types. The descriptive statistics showed that CHL learners scored close to 60% in the oral pre-test, which suggested that they could have already been quite close to the ceiling levels of the target classifiers. With such a high initial level, it might have been hard for them to improve significantly within a short period of time, regardless of which type of feedback they received during the treatment. This result was supported by the findings made of Ammar and Spada (1996). Examples 6-16 and 6-17 illustrated a typical case of a CHL learner who failed to use the target classifier even after he noticed the recasts a couple of times during the interaction.

Example 6-16 Learner XS in treatment 1
NNS: *我买了四,四,四个笔。
Wo mai le si, si, si ge bi.
1sg buy PFV four four four CL pen
I bought four pens.

NS: 那四支笔是什么颜色的？
那四支笔是什么颜色的？

Na si zhi bi shi shenme yanse de
That four CL pen is what color NOM

What color are those four pens?

NNS: 一支是绿色的，
一枝是绿色的，

Yi zhi shi lvse de
One is green, NOM

One is green,

一支是红色的。
一枝是红色的。

Yi zhi shi hongse de
One is red NOM

One is red.

Recall data
This is like zhi bi, but I was like ge bi. I didn't, or I forgot, I didn't remember what it was.

Example 6-17 Learner XS in treatment 2

NNS: *我有三个毛笔。
我有三个毛笔。

Wo you san ge maobi
I have three brushes.

NS: 我也有三支毛笔。
我也有三支毛笔。

Wo ye you san zhi maobi
I also have three brushes.

你有多少钢笔？
你有多少钢笔？

Ni you duoshao gangbi
2sg have how many pen

How many pens do you have?

NNS: 两支钢笔。
两支钢笔。

Liang zhi gangbi
Two CL pen

Two pens.

Recall data
I didn't know what gangbi was. And then the measurement word for maobi and gangbi, I didn't know.

Zhi is a frequently used classifier for long-rigid objects, which was found to emerge in L1 Mandarin Children’s vocabulary as early as age three (Hu, 1993). Therefore there
was a high likelihood that the learner XS had heard it in the past. This assumption was supported by the verbal reports in Example 6-16. The verbal reports also indicated that he noticed the target classifier provided by the NS in the recast. However, on the following day, when a different noun that required the same target classifier was presented to the learner, he failed to provide it again. As discussed in section 6.1, this could be explained as follows: since CHL learners tend to memorize a classifier and its associated noun as an inseparable linguistic unit, they may only be able to apply the classifier in the same unit, but not in a new context. It was noticeable that in both treatments, he not only noticed the target classifier, but also provided successful uptake. However, the post hoc analysis of his post-test showed that he used the general classifier *ge* instead of *zhi* in new contexts - this suggests that CHL learners need time to incorporate what they have noticed in the feedback into their internal language system. In addition, it may take some time and practice for them to produce it.

In contrast, non-CHL learners in the current study had a much lower initial level of the target classifiers (*M*<20) compared with CHL learners. Thus, it seemed reasonable to assume that meta-linguistic feedback would be more effective than recasts based on the findings made by Ammar and Spada (1996). However, although the descriptive statistics showed a slight advantage in the mean scores achieved by learners in the meta-linguistic feedback group over those in the recast group, no statistically significant difference was found in the current study. The difference between the two studies could be due to the different target structures: the target form employed in Ammar and Spada's study was the English third-person singular possessive determiners, *his* and *her*, which agree with the gender of the possessor. In contrast, the linguistic form
targeted in the current study was classifiers, which select their nouns based on different semantic domains, such as shape, function, animacy, etc. Thus the semantic relationship between a classifier and the associated noun is intricate and complex (Hu, 1993). In addition, although the target classifiers had already been encountered by non-CHL learners in the classroom, the extent to which they had been acquired varied. Some were only briefly introduced as supplementary vocabulary words; some were thoroughly taught in the text and appeared in the homework and exams. Therefore, learners might have forgotten those classifiers that were not used, as shown in Example 6-17.

Example 6-17
NNS: *我有一,一个帽子。
    Wo you yi yi ge maozi
    lsg have one one CL hat
    I have one hat.
NS: 你有一顶帽子, 是吗？
    Ni you yi ding maozi shi ma
    2sg have one CL hat is QP
    You have one hat, don’t you?
NNS: *一顶帽子。
    yi ding maozi
    one CL hat
    One hat.

*Recall data*
Yeah, I just cannot remember the measure word. But I was trying to make it more conversation. I was trying to respond or ask questions back… I wish I were more fluent. That is what I was thinking. I felt better the second day compared with the first day. Because a lot of things we discussed on the first day were things we had learned, I had studied, but I don’t ever use them. I can never talk about flowers or moon pies, or anything like that, I never talked about them. So I forget how to say them. So if you remind me, usually I can remember the majority of them.

Besides the classifiers, non-CHL learners also had to worry about their pronunciation, particularly the tones. The descriptive statistics results of learners’
perceptions showed that non-CHL learners made more phonology-related comments than CHL learners when they failed to perceive the feedback accurately. Example 6-18 presented such a case of a non-CHL learner.

Example 6-18
NNS: *我有 六花儿。
   Wo you liu huar
   1sg have six flower
   I have six flowers.
NS: 你有六枝花儿？
   Ni You liu zhi huar
   2sg have six CL flower
   You have six flowers?
NNS: 六枝花儿。
   Liu zhi huar
   six CL flower
   Six flowers.

Recall data
I just wanted to make sure that I made the correct tone for that one.

As a final explanation of lack of differences between the two feedback conditions in the current study, the length of treatment in the current study was relatively short: the total length of the two treatment tasks was approximately 20 minutes for CHL learners, and 40 minutes for the non-CHL learners. The different length of treatment for the two learner groups was due to their different fluency: the CHL learners engaged in the interaction with the NS in a smooth manner. They always responded to the NS very quickly. In contrast, the non-CHL learners tended to take some time before comprehending the NS’s question. It took them an even longer time to respond to the NS.

Previous studies that revealed significant differences between different feedback groups usually spread their treatments over a longer period of time, ranging from one hour (R. Ellis et al., 2006) to several weeks (Ammar & Spada, 2006; R. Lyster, 2004). In
fact, the descriptive statistics in the current study showed that learners in the meta-
linguistic feedback groups generally scored higher than those in the recast groups on
the post-test, despite their lower initial scores on the pre-test. This advantage could
possibly become statistically significant if learners were given a longer time to
internalize what they have increased from the feedback. However, it is not clear to what
extent the length of treatment would influence the effects of the feedback (Lownen &
Nabei, 2007). A longer treatment time may not necessarily lead to statistical differences
between different feedback groups. For instance, the academic year longitudinal study
conducted by DeKeyser (1993) found no statistical differences between implicit and
explicit feedback groups. Therefore the length of time that would be needed to optimize
the effects of different feedback needs further empirical research.

Before ending this section, I will discuss why CHL learners in the meta-linguistic
group did not outperform those in the control group on the written post-test to the extent
expected. It is also noteworthy that, although the difference between the recast group
and the control group was statistically significant, the $p$ value was .049, which was only
marginally significant.

First of all, CHL learners’ limited literacy skills might have contributed to the limited
difference between the feedback group and the control group. The written test in the
present study scored three points for each target classifier written using a correct
Chinese character, whereas two points were given for each target classifier written in
correct Hanyu Pinyin. Thus, to achieve high scores on the written test, a learner not only
needed to know the target classifier, but also the Chinese character. As explained in
Chapter 3, most HL learners only develop their literacy skills in English, but not in their
HL. The descriptive statistics showed that CHL learners scored much lower on their written test than on the oral test, indicating that even though they know a target classifier, they might not have known the Chinese character, or even the correct Hanyu Pinyin. Since feedback was provided through oral interaction, learners could not get help from the treatment in writing Chinese characters or Hanyu Pinyin.

6.3 Research Question 3

Research question 3 asked whether feedback type or language background affects the increase in learners’ performance on a) an oral imitation test, b) an untimed written cloze test.

**Oral imitation test.** The results of the current study showed that neither feedback type nor language background affected the increase in learners’ performance on the oral imitation test. In addition, the descriptive statistics showed that the CHL learners and the non-CHL learners increased by virtually the same amount when they received the same type of feedback. These results seemed to suggest that the two groups of learner benefited equally from the same type of feedback.

However, the descriptive statistics in table 5-31 showed that the CHL learners in the control group increased by nine points from the pre-test to the post-test, which was more than half of what was achieved by the CHL learners in the recast group (M=15.33), and only slightly less than half of what was scored by the CHL learners in the meta-linguistic feedback group (M=20.91). Based on this result, it is reasonable to expect that the CHL learners in the experimental groups would have increased their score even without any feedback. This could not be interpreted as practice effect, given the fact that the non-CHL learners in the control group not only did not make any
increase, but also decreased their score by 1.11 points. Apparently a more plausible interpretation is needed here.

As reviewed in Chapters 2 and 3, HL learners usually start their exposure to the HL at birth. In most of the cases they develop their HL through daily oral communication with family members. As a result, they tend to have stronger oral proficiency than FL/SL learners. In fact, HL learners were found to perform more like native speakers than the FL/SL learners on oral production tasks. The descriptive statistics in the current study, which were presented in tables 5-10 and 5-11, also showed that the CHL learners ($M > 57\%$) had a higher level in their initial knowledge of the target classifiers than the non-CHL learners ($M < 20\%$). Previous researchers attributed HL learners’ outstanding performance on oral production tasks to their implicit and automatically processed knowledge, which is typically acquired during childhood (e.g., Montrul et al., 2008). In this study, the implicit nature of HL learners’ HL knowledge was also reflected in their verbal comments collected during the stimulated recall, as illustrated in Example 6-19.

Example 6-19  A CHL learner YHX

NNS: *然后一只蛇。
Ranhou yi zhi she
Then one CL snake
Then (there is) one snake.

NS: 蛇是长长的、软软的、细细的。
She shi changchang de, ruanruan de, xixi de
Snake is long NOM soft NOM thin
Snake is long, soft, and thin.

NNS: 一条蛇。
Yi tiao she
One CL snake
One snake.

Recall data
一说出来我就知道是错的。因为我第一次想是“一条蛇”，有的时候在脑子里想的和说出来的不一样。脑子里想的是对的还是不对的，要说出来才知道。

要听到了才能知道，根据以前的经验，听过没有。

(I knew it was wrong right after I said it. Because I first thought of yitiaoshe. Sometimes what is in my mind differs from what I say. I wouldn’t know whether it is correct or not until I say it. I would only know whether it is correct or not after listening to it. According to the past experience, and according to whether I have heard it or not).

In Example 6-19, the learner immediately realized that the classifier she used was incorrect based on what she heard in the past. In other words, she could have figured out the error even without the feedback provided by the interlocutor. Therefore this case could probably be used to explain the increase obtained by the CHL learners in the control group: that is, some HL learners’ implicit knowledge could be activated through simple (repeated) exposure to the target items. For this particular learner, it is impossible to know how much of this increase was due to her implicit and intuitive knowledge developed from her past experience, and how much was due to the effect of the feedback. A post hoc analysis of YHX’s pre- and post-test showed that she increased her score from 56.3 to 89.6, which was much higher than her group average level (M=20.91). It noteworthy that her pre-test score was slightly lower than the group average (M= 57.22). Her outstanding performance could be related to her higher degree of exposure to the language compared with many of her peers in the group (see her profile in Chapter 4). Unlike many young arrivals who quickly totally shifted their dominantly language to English, YHX has kept her HL as part of her life: she speaks Mandarin with her parents, goes to Chinese Karaoki with her Chinese friends, and also watches Chinese TV programs with her parents at home. However, the feedback should have at least helped her confirm her intuitions. If her increase was mostly due to the
activation of her implicit knowledge, which type of feedback she received during the interaction should not have affected her performance.

Now I will present another learner WKW, who received recasts during the treatment. A post hoc analysis of his pre- and post-test results showed that he scored 50% on the pre-test, which was lower than the group average ($M=58.87$); and 77.1% on the post-test, which was slightly higher than the group average ($M=73.45$). His increase ($M=27.1$) was also higher than the group average ($M=15.33$). The interview data (see his profile in Chapter 4) showed that before taking a postsecondary Chinese class, his HL learning experience consisted mainly of oral communication with his parents at home. After his dominant language shifted from Mandarin to English right after age five when he started to attend mainstream schools, communication with his parents also became a mixture of Mandarin and English. Compared with YHX, his exposure to the language was more limited. Consequently, his HL knowledge was closely related to what he frequently heard from his parents during everyday life. For those classifiers that he did not hear from his parents, he usually did not have opportunities to obtain the knowledge, even those objects that were common in daily life, such as belt (see Example 6-4), mirror (Example 6-20), and hat (Example 6-21).

Example 6-20  A CHL learner WKW’s verbal comment

NNS: 也 买了 两张 镜子。
Ye mai le liangge jingzi
Also bought two mirrors.

NS: 你 买了 两 面 镜子。
Ni mai le liang mian jingzi
You bought two mirror.

NNS: 两 面 镜子。
Two mirrors.

Recall data

Jingzi, I don’t know jingzi de liangci (the classifier for mirror). Usually if it is flat, I use yizhi or yizhang. Because whenever I think of a paper, I’ll say yizhang. So whenever I am thinking something flat, I automatically assume it is yizhang. I got the impression from my mother. Normally when I do my homework, she'll point at the table, and point at the paper, and said: both of these are yizhang. She didn’t explain why.

Example 6-21 WKW’s another verbal comment

NNS: *我去商店买了一个帽子。
Wo qu shangdian mai le yi ge maozi
I go shop buy PFV one CL hat
I went to shop and bought a hat.

NS: 你那顶帽子是什么颜色的？
Ni na ding maozi shi shenme yanse de
2nd that CL hat is what color NOM
What is the color of your (that) hat?

NNS: 白色的。
Baise de
White NOM
White.

Recall data

那时候我在想很多次我跟父母去商店买衣服,可是买衣服没有什么意思。我
不知道帽子的量词。我的妈没有教我。
(At that moment I was thinking that I went shopping for clothing with my parents many time. But (buying clothing) was not fun. I don’t know the measure word for hat. My mom didn’t teach me.

In both examples 6-20 and 6-21, the learner did not know the appropriate classifier for either mirror or hat. However, a post hoc analysis of his post-test showed that he used the correct classifier for mirror, but not for hat. He did not recall that he learned either of the two classifiers from his mother. This difference could then be explained as follows: Besides being a classifier, mian (for mirror) also appears in many words which are frequently used in everyday life, such as mianbao (bread), miantiao (noodle). In
contrast, *ding*, the classifier for *hat*, is less frequently used. Therefore it could be possible that WKW associated the classifier *mian* with the words he knew (such as *mianbao*) to remember it. In the current study, both the CHL learners and the non-CHL learners reported that they tend to use this analogous strategy when acquiring an unknown classifier.

In contrast, the non-HL learners usually relied on the explicit knowledge they learned in the classroom and feedback to figure out their errors. Example 6-22 illustrates this.

Example 6-22 A non-CHL learner ZK

NNS: *我也买一张大镜字。
Wo ye mai yi zhang da jingzi
1sg also buy one CL big mirror
I also buy a big mirror.

NS: 一面很大的镜字吗？
Yi mian hen da de jingzi ma
One CL very big NOM mirror QP
A very big mirror?

NNS: 对。
Dui
Correct
Correct.

Recall data
At first I thought the measure word would be *zhang* for *jingzi*, cause it is flat surface. But it wasn’t, I was kind of surprised for a second, I heard you corrected me.

In Example 6-23, learner ZK use the rule (*zhang* is a classifier that modifies an object with a flat surface) that he learned from the class to make a reasonable assumption for *mirror*. When a recast was provided by the interlocutor, he realized that he had been corrected. His profile (see Chapter 4) showed that he had been taking Chinese courses at the university for 15 months when the data was collected. He had never been to China, and had very limited opportunities to practice his Chinese outside
the classroom. Thus his knowledge of Chinese was almost entirely developed in the classroom. Unlike YHX, he did not have rich past experiences of this language that he could activate. Therefore feedback played an important role in helping him to figure out his untargetlike language.

Lastly, it would be interesting to look closely into those who did not make any increase or even regressed from the pre-test to the post-test. A post hoc analysis of the pre- and post-test scores of these learners is presented in table 6-2.

Table 6-2. Learners who increased zero or decreased in the experimental groups.

<table>
<thead>
<tr>
<th>Learner</th>
<th>Group</th>
<th>Test type</th>
<th>Pre-test</th>
<th>Post-test</th>
<th>Increase</th>
</tr>
</thead>
<tbody>
<tr>
<td>TW</td>
<td>RHL</td>
<td>Oral</td>
<td>46</td>
<td>42</td>
<td>-4</td>
</tr>
<tr>
<td>SN</td>
<td>RHL</td>
<td>Oral</td>
<td>90</td>
<td>88</td>
<td>-2</td>
</tr>
<tr>
<td>SBY</td>
<td>RHL</td>
<td>Oral</td>
<td>92</td>
<td>88</td>
<td>-4</td>
</tr>
<tr>
<td>LL</td>
<td>MHL</td>
<td>Oral</td>
<td>90</td>
<td>88</td>
<td>-2</td>
</tr>
<tr>
<td>ZZD</td>
<td>MHL</td>
<td>Oral</td>
<td>23</td>
<td>23</td>
<td>0</td>
</tr>
<tr>
<td>LJD</td>
<td>RNHL</td>
<td>Oral</td>
<td>13</td>
<td>13</td>
<td>0</td>
</tr>
<tr>
<td>LAW</td>
<td>RNHL</td>
<td>Oral</td>
<td>25</td>
<td>21</td>
<td>-4</td>
</tr>
</tbody>
</table>

The results in table 6-2 reveal the following interesting facts: the majority of these learners (5/7) are CHL learners. Three of them (SN, SBY, and LL) were those who achieved the maximum scores on the pre-test, whereas the other two (TW and ZZD) were those who scored the lowest in the CHL learner group. In contrast, among the two non-CHL learners in the recast group, LJD scored relatively lower than the mean score, but was not among those who scored the lowest. The lowest score in the group was only six out of a possible one hundred. The final learner, LAW, scored slightly higher than the mean score (M: 19.63).

For the three CHL learners who have already achieved maximum scores on the pre-test, it would not have been not possible for them to further increase their score on the post-test. However, it is surprising that they sometimes “forgot” the accurate
classifier, which they used on the pre-test and in the treatments. This resulted in their decreased score on the post-test. Taking learner LL as an example, she used the accurate classifier *shuang* for chopsticks on her pre-test, as well as during her two treatments. However, she replaced it with an incorrect classifier *tiao* on her post-test. A similar case also occurred with learner SBY, who used the target classifier *pi* for horse on his pre-test and the two treatments, but replaced it with the incorrect classifier *tiao* on the post-test. It might not be a coincidence that both *shuang* and *pi* were the classifiers appeared in the original recorded sentences on the post-test, in which learners were expected to modify the original sentence if necessary. A plausible explanation for this type of “regression” is that, as discussed earlier in Chapter 6, CHL learners rely heavily on their past listening experience when dealing with classifiers, particularly when they have rich exposure to the HL at home. However, these experiences are not always accurate. Sometimes learners assume an incorrect classifier as an alternative one (see Example 6-6). Therefore for LL and SBY, it is possible that in the past they had heard both the target classifier and the one used in the original recorded sentences on the post-test, and assumed that they were interchangeable.

The above explanation is further evidenced by the fact that the same type of “regression” was not seen among the two non-CHL learners, as well as learner ZZD, who scored the lowest in the CHL learner group. The two non-CHL learners’ “lack of progress” was mostly caused by their reduced usage of the general classifier *ge* on their post-test. Although this resulted in the decrease of their score, it showed the improvement in their awareness of specific classifiers. On the other hand, the CHL learner ZZD’s “no progress” on the post-test was mostly due to his incorrect choice of
specific classifiers. For instance, he used *zhang* for dragon, *tiao* for gloves on the post-test. The interview data reveals that ZZD had much more limited exposure to Mandarin compared with SN, SBY, and LL: as the third born of the family, and with parents who speak fluent English, ZZD communicates with his family members mostly in English. Although he attended Chinese language school once a week between age eight and ten, his formal education in Mandarin at the postsecondary level (*n*=1 month) was much shorter compared with the non-CHL learners in the current study. In other words, he did not have as wide a range of classifiers as learners SN, SBY, and LL had on the one hand; he also did not have some explicit knowledge that the non-CHL learners possessed. Most of the non-CHL learners in the current study knew that *zhang* should be used for an object that is flat or has a flat surface. None of them used *zhang* to classify a dragon.

Unlike ZZD, the other CHL learner TW displayed a “regression” that was similar to that of SBY and LL: he used the accurate classifier *shuang* for chopsticks on his pre-test, as well as during his two treatments, but replaced it with an incorrect classifier *tiao* on his post-test. Table 6-3 summarizes the profile of TW and SBY, who both decreased the most from the pre-test to the post-test among the CHL learners. Despite the similarities in their generation, age of arrival, and schooling in the home country, which are considered critical predictive variables of learners’ proficiency in their HL, the two learners scored very differently on their tests: one scored the highest in the group, the other was one who scored the lowest. The most distinctive difference that set these two learners apart is their home language environment: TW’s home language was English dominant, whereas SBY’s home language was Mandarin dominant. As a result, TW
scored lower than his CHL peers, and could not activate what he acquired by age ten within a two-day treatment, in spite of his relatively older age of arrival and much longer formal schooling in the home country compared with other CHL learners in the current study,

In summary, the CHL learners developed a rich implicit knowledge of classifiers due to their early exposure to the language, which may not always be visible in their oral production. However, with some assistance (such as reoccurrence in the pre- and post-test), some of this knowledge could be activated, as we saw among the CHL learners in the control group. Being facilitated by feedback, the CHL learners made even greater increases on the oral post-test. However, since their initial levels of implicit knowledge have already approximated ceiling levels, although meta-linguistic feedback still outperformed recasts, its superiority was not significant. Those who have scored the maximum even showed slight regression on the post-test.

Table 6-3. The profile of learners TW and SBY

<table>
<thead>
<tr>
<th></th>
<th>TW</th>
<th>SBY</th>
</tr>
</thead>
<tbody>
<tr>
<td>Generation</td>
<td>First</td>
<td>First</td>
</tr>
<tr>
<td>Age of arrival</td>
<td>Ten</td>
<td>Ten</td>
</tr>
<tr>
<td>Schooling in home country</td>
<td>Four years</td>
<td>Four years</td>
</tr>
<tr>
<td>Cultural identity</td>
<td>Less stronger</td>
<td>Stronger</td>
</tr>
<tr>
<td>Home language</td>
<td>Step father: native English speaker</td>
<td>Father: Mandarin/ English Mother: Mandarin/ English</td>
</tr>
<tr>
<td></td>
<td>Mother: English/ Mandarin</td>
<td></td>
</tr>
<tr>
<td>Home literacy environment</td>
<td>Not so many</td>
<td>Many</td>
</tr>
</tbody>
</table>

Written cloze test. In contrast to findings from the oral test, the results of the written tests showed that feedback type significantly affected learners’ increase in their performance from the pre-test to the post-test. That is, meta-linguistic feedback led to a greater increase than recasts. However, this superiority was found only among the non-CHL learners, but not among the CHL learners. That is, the scores of the non-CHL
learners in the meta-linguistic feedback group \((M=21.92)\) increased significantly more than those of the non-CHL learners in the recast group \((M=9.17)\). In contrast, the CHL learners in the two experimental groups increased similarly \((15.33 \text{ vs. } 16.45)\).

As discussed earlier in this chapter, recasts were more explicit in the current study than in some previous studies due to the laboratory setting. Consequently both recast and meta-linguistic feedback were generally perceived as overtly corrective by learners. However, meta-linguistic feedback indicated the source of learners’ error with meta-linguistic information, which can help learners obtain rule-based knowledge. In contrast, recasts simply reformulated learners’ problematic utterance without any explanation. Unlike the CHL learners who had intuitive and implicit knowledge, the non-CHL learners mostly relied on the feedback to figure out the gap between their utterance and the target linguistic forms. Thus the feedback type had a greater impact on the non-CHL learners’ test performance than the CHL learners, particularly when the target linguistic form was relatively complex, as shown in examples 6-23 and 6-24.

The two classifiers for flowers were found to be a challenge among the participants of the current study, particularly the non-CHL learners. In Example 6-23, the interlocutor clearly indicated that flowers that grow on the same stem should not be classified by *duo*. The learner’s verbal comments reported that she understood the interlocutor’s explanation. In contrast, in Example 6-24, the interlocutor only embedded the target classifier in her recast. The learner reported that he was still confused about these two classifiers even after the two-day treatment. A post hoc analysis also showed that the learner in Example 6-23 used the correct classifier on the written post-test, whereas the learner in Example 6-24 did not.
Example 6-23 A non-CHL learner in the meta-linguistic feedback group

NS: 你买了多少花儿？
    Ni mai le duoshao huar
    2sg buy PFV how many flower
    How many flowers did you buy?

NNS: 五朵花。
    Wu duo huar.
    Five CL flower
    Five flowers.

NS: 这五朵花儿都在一个枝子上，这样的我们不可以五朵花。
    (This five flowers all grow in the same stem. We can’t call them duo)
NNS: ……

NS: 一枝花。
    Yi zhi hua
    One CL flower
    One flower.

NNS: 一枝花。
    Yi zhi hua
    One CL flower
    One flower.

Recall data
I was processing, putting into my head, zhi means stick-like thing. After you told me, I was like oh, ok, so I should remember zhi for things that are like sticks.

Example 6-24 A non-CHL learner in the recast group

NS: 你那枝白色的花儿上有多少花儿？
    Ni na zhi baise de huar shang you duoshao huar
    2sg that CL white NOM flower on have how many flower
    How many flowers are on that white bunch of flowers?

NNS: 有五枝花儿。
    You wu zhi huar
    Have five CL flower
    (There are) five flowers.

NS: 有五朵花儿。
    You wu duo huar
    Have five CL flower
    (There are) five flowers.

NNS: 五朵花儿。
    You wu duo huar
    Have five CL flower
    (There are) five flowers.
Recall data
Did you tell me different measure words, one of them was for flowers, one of them was for the individual flowers, or something? I was confused right there, cause I was trying to figure out. I think there was one flower, or one plant with a bunch of flowers. I was trying to say how many flowers were on that plant, but it sounded like I had a bunch of flowers. So I guess I was thinking how am I goanna say this, I am goanna to say the differences between those two things.

It is noteworthy that the learner in Example 6-23 did not use the accurate classifier on the oral post-test. A possible explanation for this result is that the learner did not have time to draw on his newly acquired explicit knowledge under the time pressure. In fact, the non-CHL learners generally scored higher on the written test than on the oral test. If explicit knowledge is most useful when learners have time to draw on it, then it is not surprising that the non-CHL learners performed better on the untimed written test than the oral imitation test. In contrast, the CHL learners scored lower on their written test than oral test, which also went along with the discussion presented earlier. That is, these learners had limited explicit knowledge compared with their rich implicit knowledge.

Besides explicit knowledge, learners’ performance in the current study was closely related to their knowledge of Chinese characters, as a full mark was given to each target classifier only when it was written in Chinese characters. As reviewed in Chapter 2, HL learners usually develop their HL literacy skills much later than their oral proficiency. Many of them even remain illiterate for the rest of their life. Previous research did not find that CHL learners had any advantages over non-CHL learners in their Chinese character learning (Ke, 1998; Xiao, 2006). In the current study, the average length of postsecondary level formal classroom instruction of Chinese was 10.9
months for the CHL learners, but 19.4 months for the non-CHL learners. Table 6-4 summarizes learners’ use of Chinese characters on the written pre- and post-test.

The results in Table 6-4 indicate that with the exception of the CHL learners in the recast group, learners generally show a similar level of knowledge of Chinese characters. The non-CHL learners used even more Chinese characters than the CHL learners. Although the non-CHL learners in the two experimental groups showed a similar level of knowledge of Chinese characters, those in the meta-linguistic feedback group outperformed those in the recast group by approximately eight percent. A possible explanation for this is that the meta-linguistic information in the feedback helped learners connect the target classifier and the associated Chinese character which they had acquired in other contexts in the past, which is illustrated in Example 6-25.

Table 6-4. Learners’ knowledge of Chinese characters on the pre- and post-test

<table>
<thead>
<tr>
<th></th>
<th>CHL learners</th>
<th></th>
<th>Non-CHL learners</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Recast</td>
<td>Meta</td>
<td>Recast</td>
<td>Meta</td>
</tr>
<tr>
<td>Pre-test</td>
<td>136 (51.5%)</td>
<td>74 (30.5%)</td>
<td>159 (60.7%)</td>
<td>148 (56.0%)</td>
</tr>
<tr>
<td>Post-test</td>
<td>78 (29.5%)</td>
<td>40 (16.5%)</td>
<td>94 (35.6%)</td>
<td>77 (29.2%)</td>
</tr>
<tr>
<td>Pre-test (without ge)</td>
<td>77 (29.2%)</td>
<td>34 (14.0%)</td>
<td>51 (19.3%)</td>
<td>48 (18.2%)</td>
</tr>
<tr>
<td>Post-test (without ge)</td>
<td>72 (27.3%)</td>
<td>37 (15.3%)</td>
<td>54 (20.5%)</td>
<td>74 (28.0%)</td>
</tr>
</tbody>
</table>

Example 6-25 A non-CHL learner in the recast group
NNS: *一件帽子。
Yi jian maozi
One hat.

NS: 帽子的话虽然我们是 on top。 但是我们有另一个. 这个我们叫什么？
You wu duo huar
(There are) five flowers.

NNS:  …
NS: Top of the head 我们叫什么？Top of the mountain？
NNS：一顶帽子。
Recall data

That helps. I will never forget that after you said *touding*. Like the mountain, the top of the mountain. I remembered from that *ding* would be the top and peak, so. That was good.

6.4 Implications

6.4.1 Theoretical Implications

The main purpose of the current study was to compare the relative effects of explicit and implicit feedback by empirically investigating learners with heterogeneous language backgrounds and learning experiences. Therefore the findings of the study particularly extended our knowledge of previous findings that drew largely from EFL/ESL learners who had homogeneous language backgrounds and learning experiences.

Previous research generally found that both explicit and implicit feedback facilitated language acquisition. However, in terms of the relative effects of implicit and explicit feedback, the results were mixed. Generally, studies conducted in laboratory and CALL settings found no significant differences between implicit and explicit feedback (e.g., Carroll & Swain, 1993; Kang, 2009; Kim & Mathes, 2001; Leeman, 2003; Nagata, 1993; Sauro, 2009). On the other hand, studies conducted in classroom settings revealed more complicated results: some found no significant differences between implicit feedback and explicit feedback (e.g., DeKeyser, 1993; Loewen & Nabei, 2007; Kim & Mathes, 2001; Lyster & Izquierdo, 2009); while others showed that feedback in a more explicit form (e.g., prompts) worked more efficiently than feedback in a more implicit form (e.g., recasts) (e.g., Ammar & Spada, 2006; Havranek & Cesnik, 2001; Lyster, 2004; Muranoi, 2000; Sheen, 2007; R. Ellis et al., 2006). The contrasting results found in the previous studies suggested that further investigation into the relative effects of implicit and explicit feedback was necessary.
In the investigation of the relative effects of implicit and explicit feedback, the current study, overall, did not find a significant difference between explicit feedback in the form of meta-linguistic feedback and implicit feedback in the form of recasts. Although this result was typical in laboratory studies, the current study expanded the generalizability of previous findings particularly by including the following two areas that had not been addressed in previous studies: (1) learners’ language background, and (2) learners’ perceptions of implicit and explicit feedback.

Learners recruited in previous studies were usually FL or SL learners, who only received formal language instruction in classrooms. As a result, they had higher levels of explicit knowledge than implicit knowledge. Although extensive research has been conducted to investigate the effects of implicit feedback in the form of recasts and explicit feedback in the form of meta-linguistic feedback, few of these studies has recruited learners who had a higher level of implicit knowledge than explicit knowledge due to their exposure to the HL during the childhood. Kang was one of the few researchers who recruited HL learners; however he did not collect these learners’ background information and include it in the discussion of his results (2009). Ammar and Spada (2006) addressed the relationship between learners’ proficiency levels and the effects of implicit and explicit feedback, and found that high-proficiency learners benefited equally from both types of feedback. In contrast, low-proficiency learners benefited more from explicit feedback in the form of prompts (including meta-linguistic feedback, repetition, and elicitation) than from recasts. However, their learner population had no HL background. The current study filled the gap in the literature by investigating HL learners in comparison with learners without HL background. In
addition, to the best of my knowledge, it was the first research that targeted CHL learners.

Drawing from findings of previous research on HL learners’ perception of feedback, the current study also found that CHL learners perceived feedback more accurately than non-CHL learners (Gass & Lewis, 2007). In addition, the current study took a further step to explore feedback types and learners’ perception, which was not investigated in the previous research. The results showed that CHL learners significantly outperformed non-CHL learners in accurately perceiving meta-linguistic feedback. This result suggested that meta-linguistic feedback may be more effective in drawing learners’ attention temporarily from meaning to form during task-based interaction, particularly when learners are more meaning-oriented, and possess a high level of implicit knowledge, but lack explicit knowledge to help them notice the gap between their interlanguage and the target language, as was the case with the CHL learners in the current study.

Attention that focuses on forms, as well as noticing of the gap (Schmidt & Frota, 1986) is considered a step toward change in learners’ language knowledge. However, for the actual change to eventually occur and show up in learner’s performance, a certain amount of time is needed for learners to incorporate the newly acquired knowledge into their language system (Mackey et al., 2000). This argument may explain why meta-linguistic feedback did not bring a significant advantage over recasts in the increase of knowledge made by CHL learners in the post-test, although it helped them to notice more discrepancies between their non-targetlike forms and the target-like forms.
6.4.2 Pedagogical Implications

The participants in the current study included a combination of traditional FL learners of Chinese and learners who had early exposure to Chinese as a HL. They represent the diverse learner population in postsecondary Chinese programs in the United States. Therefore the findings in the current study provide invaluable implications for Chinese language education, particularly in language curriculum development and classroom instruction.

Currently, postsecondary Chinese language classes generally implement either a single-track or a dual-track language program. The former puts all the learners into the same curriculum, ignoring their language background; the latter separates learners with CHL background from those without CHL background. A dual-track language program usually includes two types of classes: the regular class, in which CHL learners and non-CHL learners attend the same class; and the HL class, in which CHL learners with strong oral speaking proficiency and some literacy skills are enrolled. In either track, the curriculum is designed based on an assumption that the traditional speaking, listening, reading and writing skills should be equally emphasized. The participants of the current study were recruited from these two types of classes. However, the results showed that the CHL learners had a much higher level of oral proficiency than literacy skills. In contrast, the non-CHL learners demonstrated more evenly distributed oral and literacy skills. Thus the current curriculum is apparently more able to accommodate the language needs of non-CHL learners than CHL learners. The results of the current study showed a great gap between the oral and written skills presented by the CHL learners. In addition, the CHL learners scored very close to the non-CHL learners on their written test. Therefore although CHL learners still need to improve their oral productive skills in
certain areas, such as vocabulary, register, etc., they should identify writing skills development as a priority. A new curriculum that can better satisfy the different learners’ language needs is strongly needed in the field of Chinese language instruction.

Besides the language curriculum, Chinese language instructors also need to adjust their classroom instruction to maximize the learning effects. The results of the current study showed that the CHL learners tended to rely on what they have heard in the past when assessing the accuracy of their language, instead of using rule-based grammatical knowledge. In addition, they had a natural tendency to focus their attention more on content rather than linguistic forms when engaging in a conversational interaction. As claimed in the Noticing Hypothesis, learners’ noticing of the gap between their internal language system and the target language system is the precondition for converting input to intake (Schmidt, 1990, 1993, 1994, 1995), which can further lead to a change in their interlanguage. Therefore, one of the most critical goals in Chinese classroom instruction should be facilitating CHL learners to notice the discrepancies between their problematic output and the targetlike input provided by the instructor. As CHL learners normally acquire their HL implicitly (at least partially), they tend to lack explicit meta- and rule-based knowledge that is needed in order to identify the gap. Therefore feedback that explicitly indicates learners’ errors, such as meta-linguistic feedback employed in the current study, can be an efficient tool when the instructor tries to draw CHL learners’ attention from meaning to certain linguistic forms, particularly for CHL learners’ stabilized non-target like language.

On the other hand, non-CHL learners develop their Chinese knowledge mainly through explicit learning in a classroom setting. Therefore they develop explicit
knowledge and literacy skills before implicit knowledge. These learners need to convert their explicit knowledge into implicit knowledge in order to become more fluent in their oral production. The explicit feedback in the context of a communicative task can facilitate the conversion from explicit knowledge into implicit knowledge. For this reason, implicit feedback such as recasts could provide non-HL learners more intentional practice.

Lastly, the results of the current study also suggested that both CHL learners and non-CHL learners need to improve their literacy skills, particularly where Chinese characters are concerned. While the oral corrective feedback, particularly explicit feedback, was found to help the non-CHL learners to relate the Chinese characters that they acquired in the past to new linguistic contexts, it did not work equally efficiently for the CHL learners. Based on these findings, Chinese instructors may find it helpful to provide their feedback in a written form in the instruction of Chinese characters.

6.5 Limitations and Future Research

Having discussed the potential contributions of the findings to the field of SLA and HL instruction theoretically and pedagogically, I will point out some limitations of the current study from the following three aspects: the generalizability of the current findings, the design, and the methodology.

First of all, the current study was conducted in a tightly controlled laboratory setting, in which the explicitness of corrective feedback and the linguistic forms were tightly controlled. In addition, the feedback was delivered intensively on the target linguistic forms in dyadic interactions between an individual participant and the NS interlocutor on a one-to-one manner. All these factors might have weakened the ecological validity of the current findings. Thus the implications of the study may not
simply expand to studies conducted in classroom settings, where learners are exposed to mixed types of feedback on various linguistic forms (Lyster & Ranta, 1997; Sheen, 2006). Another limitation that might weaken the generalizability of the current findings concerns the participants of the study. Due to the difficulties of recruiting adequate numbers of participants, the current study only included CHL learners who grew up in a Mandarin-speaking family. Learners who grew up in a home where a Chinese dialect other than Mandarin was spoken (e.g., Cantonese, Shanghainese, etc), were excluded from the current study. It is not clear whether CHL learners with different dialect backgrounds would benefit from implicit and explicit feedback differently. In addition, the sample size of the current study \((N = 64)\) was relatively small, due to the difficulty of recruiting a large number of participants at the intermediate level and above for a less commonly taught language. Lastly, due to the small participant population, an experimental group that engages only in an interaction without receiving any type of feedback was not included in the current study. Without a comparison with this interaction-only group, it was unclear whether the improvements learners made from the pre-test to the post-test was entirely due to the feedback, or partially due to interaction only. The CHL learners are particularly in question concerning this issue given the fact that scores of the CHL learners in the control group increased considerably from the pre-test to the post-test, particularly on their oral test. Further studies should include an additional control group to allow for further comparison.

Secondly, in terms of the design shortcoming, the current study did not employ a delayed post-test due to the difficulty of recruiting adequate subjects who were willing to commit one more day several weeks apart from the three-day activity. A delayed post-
test was employed by many researchers to examine the effects of implicit and explicit feedback in the past (e.g., Ammar & Spada, 2006; Carroll & Swain, Kang, 2009; Leeman, 2003; R. Ellis et al., 2006; Sheen, 2007). R. Ellis and his colleagues did not find that significant differences existed between implicit and explicit feedback groups in the immediate post-test; however, they found them in the delayed post-test (R. Ellis et al., 2006). Sheen (2007) found that the meta-linguistic feedback group significantly outperformed the recast group on both immediate and delayed post-tests. Previous research found that it takes time for learners to incorporate the target structure into their interlanguage systems, thus the effects of feedback become more evident on in the delayed test, particularly when the effects were measured by tests that favored implicit knowledge (Mackey, 1999, also see R. Ellis et al., 2006). Due to the lack of a delayed post-test, learners’ measurements of improvements of target forms in the current study were limited to only short-term effects, as shown on the immediate post-test, thus potentially weakening the findings of research questions 2 and 3.

A final shortcoming of the test design was the inconsistency between the oral and written test in terms of identifying non-targetlike classifiers. On the written test, learners were required to write in Chinese characters as much as they could, so that when a learner used a homonym of the target classifier it could be identified. However, the same type of error could not be identified on the oral test. This limitation particularly favored the CHL learners on their oral test, as they acquired their HL mainly through listening to their parents. In many cases, they only acquired the oral form rather than the meaning of a classifier. For instance, the CHL learners’ written test results showed that they often misused ฤ ฤ zhi, which is the classifier for animals, to classify flowers, which
is supposed to be 枝 zhi. A possible solution to solve the inconsistency between the oral and the written test in future studies would be to require the learner to orally describe to which zhi he/ she is referring, although this would detract from the naturalness of the interaction..

Finally, I will discuss a methodological limitation in collecting learners’ noticing of feedback. The current study employed stimulated recall, which has been widely used in previous studies to elicit introspective verbal comments about feedback (e.g., Adams, 2003; Egi, 2007; Gass & Lewis, 2007; Mackey et al., 2000; Swain & Lapkin, 2002). However, the method has some limitations: first of all, verbal comments collected through stimulated recall present only a part of learners’ cognitive processes. Learners may not provide verbal comments even though they have noticed feedback. They also may only report one type of perception about feedback when they actually perceive it in several ways (Egi, 2004). Lastly, there were two shortcomings in processing the stimulated recall in the current study. The first one was that the stimulated recall was not conducted immediately after the treatment. The second one was that the stimulated recall and the treatment were both conducted by the researcher. As always, further research is needed.

6.6 Conclusion

The present study empirically investigated the effects of implicit feedback in the form of recasts and explicit feedback in the form of meta-linguistic feedback on Chinese language learners’ acquisition of classifiers. The results showed that both feedback types were effective in facilitating learners’ perception of the corrective intention of feedback. In addition, both feedback groups showed significant increase relative to the
control group on the post-test, indicating the overall effectiveness of meta-linguistic feedback and recasts on learners’ interlanguage, regardless of their language background.

On the other hand, the study also found that learners’ language background may have some impacts on the effects of meta-linguistic feedback and recasts. The study examined the relative effectiveness of meta-linguistic feedback and recasts on an unexplored learner group: CHL learners who acquired Chinese mainly through implicit learning at home and in the Chinese community, with only a short time of formal education at the postsecondary level; in comparison with non-CHL learners who acquired Chinese mainly through explicit learning in the classroom. The results shed some light on how the relative effectiveness of feedback may be affected by learners’ language background, particularly with the following findings: a) the CHL learners in the explicit feedback condition perceived feedback more accurately than the non-CHL learners in the implicit feedback condition; b) explicit feedback proved superior to recasts in promoting the non-CHL learners’ writing skills ($p=.002$). However, the same superiority was not seen for the CHL learners.

However, due to the limitations in design and methodology, further research is needed to increase our understanding of how CHL learners acquire or require their CHL.

---

1 The transcription remained the original comments made by learners without modifying their grammatical errors.
2 The transcription remained the original comments made by learners without modifying their grammatical errors.
3 This may not apply to all the CHL learners as some of them skip the first year of the Chinese course.
4 The general classifier 个 ge was the most frequently used Chinese character on the written post-test.
APPENDIX A
TRANSCRIPTION CONVENTIONS

CL  Classifier
CP  Compliment
CRS Currently relevant state (le)
NOM nominalizer
PFV perfective aspect (-le)
Pre preposition
QP question particle
1sg First person singular
2sg Second person singular
3sg Third person singular
1pl First person plural
APPENDIX B
CONSENT FORM

I.R.B No.: 2009-U-0172

Informed Consent

Protocol Title: The effects of implicit and explicit feedback
Please read this consent document carefully before you decide to participate in this study.

Purpose of the research study: The purpose of this study is to examine the effects of feedback techniques on heritage Chinese language learners.

What you will be asked to do in the study: You will participate in three-days of activities: Day 1: You will be interviewed by the researcher about your Chinese heritage language learning experience at home (approximately 30-40 minutes). The interview will be audio-taped. You will also complete a set of written and oral questions (five minutes for each) in Chinese, the purpose of which is to measure your current proficiency level in Chinese; Day 2: You will be asked to tell a story based on provided pictures. In between, the researcher may ask you some questions related to your story (approximately 40 minutes). The process will be both video and audio taped. You will also complete a set of written and oral questions in Chinese in order to confirm your proficiency level in Chinese. Day 3: You will be shown some video clips chosen from the interview of Day 1, and will be asked to give some comments about them (approximately one hour). Your participation in the study has no impact on your school grade, and your results of written and oral questions will not be communicated to your instructors. The process will be audio taped. All the video and audio recordings will only be used for data analysis purposes.

Time required: 3-4 hours

Risks and Benefits: You will be practicing Chinese with a native speaker by participating in the study. No risks will be involved in the study.

Compensation: No compensation will be given for participating in this research.

Confidentiality: Your identity will be kept confidential to the extent provided by law. Your information will be assigned a code number. The list connecting your name to this number will be kept in a locked file in my faculty supervisor's office. When the study is completed and the data have been analyzed, the list will be destroyed. Your name will not be used in any report.

Voluntary participation: Your participation in this study is completely voluntary. There is no penalty for not participating.

Right to withdraw from the study: You have the right to withdraw from the study at anytime without consequence.
Whom to contact if you have questions about the study:

Han Ye, Graduate Student, Linguistics Program, phone: 352-262-8841. Email: hanye@ufl.edu, 4131 Turlington Hall, PO Box 115454, Gainesville FL 32611-5454.

Antes, Theresa. Professor, Romance Languages and Literatures, phone: (352) 392-2101 x 236, Email: antes@rll.ufl.edu, Dauer Hall PO BOX 115565, Gainesville, Florida

Whom to contact about your rights as a research participant in the study: IRB02 Office, Box 112250, University of Florida, Gainesville, FL 32611-2250; phone 392-0433.

Agreement: I have read the procedure described above. I voluntarily agree to participate in the procedure and I have received a copy of this description.

Participant: _____________________________ Date: _________________

Principal Investigator: _____________________ Date: _________________
APPENDIX C
QUESTIONNAIRE

1. Name ________________________________________________ (both your Chinese and English name)

2. Gender:   F     M

3. Class Standing: Freshman/Sophomore/ Junior/ Senior/ Graduate

4. Your Chinese professor’s name ____________ 5. Your email address _____________________


7. If your answer is No in No.4, your age of immigration _________________

   Did you migrate from the following regions? Please circle one: Mainland China Taiwan   Hong Kong

8. What was the first language you spoke? English    Mandarin   Cantonese   Others _______________

9. Which language do you speak with your parents NOW (Choose all the languages you speak with them)?
   English   Mandarin   Cantonese   Others _______________

10. Which language do you speak with your siblings NOW (Choose all the languages you speak with them)?
    English   Mandarin   Cantonese   Others _______________

11. Which language do your parents speak with each other (Choose all the languages they speak)?
    English   Mandarin   Cantonese   Others _______________

12. Which language do you speak with your friends who can speak Chinese NOW?
    Mandarin   Cantonese   Others _______________    English _______________

13. Which language do you speak with your grand parents?
    English   Mandarin   Cantonese   Others _______________
14. Which language did your parents teach you how to write at home?

Mandarin  Cantonese  English _______________  Others _______________
APPENDIX D
WRITTEN PRE-TEST

Please complete the following sentences. You can write either in Chinese characters or in Pin Yin.

1. 我有_________和_________。
   I have three cats and two horses.
2. _______我吃了_________。
   I ate a chicken for dinner.
3. 妹妹__________________。
   My little sister drew a dragon.
4. _______的宠物是_________。
   My little brother's pet is a snake.
5. 桌子上放着_________和_________。
   There are a pair of chopsticks and two pairs of scissors on the table.
6. 明天是我的_______，妈妈给我买了_________。
   Tomorrow is my birthday, my mom bought a pair of shoes for me.
7. 昨天我去_________的时候，丢了_________。
   I lost one of my gloves yesterday when I went to watch a movie.
8. 他这个人真_______，脚上只穿了_________鞋。
   He is really weird, he only wears one shoe on his foot.
9. 她的包里有_________和_________。
   There is one toothbrush and one pen in her bag.
10. 我昨天去_________买了很多东西：_________、_________、_________和_________。
    Yesterday I went to a shopping center and bought a lot of stuff: two shirts, one sweater, and one belt.
11. 小高家的_________里有_________。
    There is a piano in Xiao Gao's living room.
12. 我花了一万五千_________买了_________。
    I spent $15000 on buying a car.
13. _______是爸爸的生日，我想送给他_________和_________。
    The day after tomorrow is my dad’s birthday, I want to give him a tie and a hat.
14. 我早上十点才_________，早饭只吃了_________和_________。
    I didn't get up until 10 o'clock. I only had one slice of bread and three grapes.
15. 情人节那天，我送给我女朋友_________红玫瑰。
    On Valentine's day, I gave by girlfriend eighteen red roses.
16. 我的手链上有_________珍珠。
    There are eleven pearls in my bracelet.
17. 小白的_________里有_________。
There is a mirror in Xiao Bai’s living room.

18. 那天我吃了________________。

19. 那天我吃了________________。

19. 教室里有__________________。

20. 教室里有__________________。

20. 桌子上的盘子里有__________________。

21. 桌子上的盘子里有__________________。

21. 盒子里有__________________和__________________。

22. 盒子里有__________________和__________________。

22. 姐姐想学十字绣，就去__________________买了__________________。

My sister wants to learn cross-stitch, so she went to the shop and bought three sewing needles.
APPENDIX E
WRITTEN POST-TEST

Please complete the following sentences. You can write either in Chinese characters or in Pin Yin.

1. ______的宠物是______________。
   My little brother's pet is a snake.
2. 我有______________和______________。
   I have three cats and two horses.
3. ____________我吃了______________。
   I ate a chicken for dinner.
4. 姐姐想学十字绣, 就去______________买了______________。
   My sister wants to learn cross-stitch, so she went to the shop and bought three sewing needles.
5. 教室里有________________________。
   There are twenty-three tables in the classroom.
6. 昨天我去__________的时候, 丢了______________。
   I lost one of my gloves yesterday when I went to watch a movie.
7. 我早上十点才__________，早饭只吃了______________和______________。
   I didn't get up until 10 o'clock. I only had one slice of bread and three grapes.
8. 小白的__________裏有________________________。
   There is a mirror in Xiao Bai's living room.
9. 她的包裏有______________和______________。
   There is one toothbrush and one pen in her bag.
10. ____________桌子上的盘子里有________________________。
    There is beef on the plate on the table in the kitchen
11. ____________桌子上放着______________和______________。
    There are a pair of chopsticks and two pairs of scissors on the table.
12. 我昨天去__________买了很多东西: ____________、______________和______________。
    Yesterday I went to a shopping center and bought a lot of stuff: two shirts, one sweater, and one belt.
13. 盒子里有______________和______________。
    There are five leaves and five flowers in the box.
14. 小高家的________________________里有________________________。
    There is a piano in Xiao Gao's living room.
15. 妹妹________________________。
    My little sister drew a dragon.
16. ________ is爸爸的生日, 我想送给他________________________和
________________________。

The day after tomorrow is my dad’s birthday, I want to give him a tie and a hat.

17. 情人节那天，我送给女朋友________________________红玫瑰。

On Valentine’s day, I gave my girlfriend eighteen red roses.

18. 我的手链上有________________珍珠。

There are eleven pearls in my bracelet.

19. 明天是我的________，妈妈给我买了________。

Tomorrow is my birthday, my mom bought a pair of shoes for me.

20. ________那天我吃了________________。

At the Mid-Autumn festival, I ate two mooncakes.

21. 我花了一万五千______买了________________。

I spent $15000 on buying a car.

22. 他这个人真________，脚上只穿了________鞋。

He is really weird, he only wears one shoe on his foot.
APPENDIX F
TREATMENT 1

Tell me what you bought
APPENDIX G
TREATMENT 2
LIST OF REFERENCES


In S. Foster-Cohen & A. Nizegorodzew (Eds.), EUROSLA Yearbook (pp. Amsterdam: Benjamins.
Pragmatics, 14, 199-216.
Xiao (Eds.), Chinese as a heritage language: Fostering rooted world citizenry
(pp. 1-12). Honolulu: University of Hawai'i Press.
heritage language. In A. W. He & Y. Xiao (Eds.), Chinese as a heritage
language: Fostering rooted world citizenry (pp. 1-12). Honolulu: University of
Hawai'i Press.
Qualitative research in applied linguistics: A practical introduction (pp. 91-111).
Hampshire: Palgrave Macmillan.
Hendrys, J. D. (2008). The Chinese heritage language learners’ existing linguistic
knowledge and abilities. In A. W. He & Y. Xiao (Eds.), Chinese as a heritage
language: Fostering rooted world citizenry (pp. 53-66). Honolulu: University of
Hawai'i Press.
Hendryx, J. D. (2008). The Chinese heritage language learners’ existing linguistic
knowledge and abilities. In A. W. He & Y. Xiao (Eds.), Chinese as a heritage
language: Fostering rooted world citizenry (pp. 53-66). Honolulu: University of
Hawai'i Press.
linguistic autobiographies. ERIC Digest, Retrieved March 18, 2010, from
Huang, s. (1993). Yuyan, shehui yu zuqun yishi. [Language, society, and ethnicity].
Taipei: Crane.
recent Chinese immigrants in New York City. In A. W. He & Y. Xiao (Eds.),
Chinese as a heritage language: Fostering rooted world citizenry (pp. 189-203).
Honolulu: University of Hawai'i Press.


BIOGRAPHICAL SKETCH

Ye Han was born in Beijing, China. She received her B.A. in Teaching Chinese as a Second Language from Beijing Language and Culture University in China. She received her M.A. in Journalism from Sophia University in Tokyo Japan. She received her Ph. D in Linguistics from the University of Florida in the summer of 2010.