3
2010

LIBRARY
Michigan State
University

This is to certify that the
dissertation entitled

THE NOTICING AND EFFECT OF TEACHER FEEDBACK IN ESL CLASSROOMS

presented by

WEIQING WANG

has been accepted towards fulfillment
of the requirements for the

Ph.D. degree in Curriculum, Teaching, and Educational Policy

Major Professor's Signature

16 Dec. 2009

Date

MSU is an Affirmative Action/Equal Opportunity Employer

PLACE IN RETURN BOX to remove this checkout from your record.
TO AVOID FINES return on or before date due.
MAY BE RECALLED with earlier due date if requested.

DATE DUE	DATE DUE	DATE DUE

5/08 K:/Proj/Acc&Pres/CIRC/DateDue.indd

THE NOTICING AND EFFECT OF TEACHER FEEDBACK IN ESL CLASSROOMS

By

Weiqing Wang

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Curriculum, Teaching, and Educational Policy

2009

ABSTRACT

THE NOTICING AND EFFECT OF TEACHER FEEDBACK IN ESL CLASSROOMS

By

Weiqing Wang

This study examined learners' noticing of teacher feedback and their language improvement from three aspects: the characteristics of feedback episodes, the nonlinguistic cues in teacher feedback, and the metalanguage in teacher feedback. The data came from classroom observations, stimulated recall interviews, and two tailor-made individualized tests: an immediate test and a delayed test.

A total of nine classes were observed, with four to six observations in each class. Lessons were both audio and video taped. Both the immediate test and stimulated recall interview were carried out one to three days after the observation(s) from which they were developed while the delayed test was administered 14 to 16 days after the observation(s) from which it was developed. Altogether, there were 1434 feedback episodes, 279 stimulated recall comments, 272 test items in the immediate test, and 282 test items in the delayed test.

The results suggest that all the seven characteristics of feedback episodes examined had a statistically significant effect on both the occurrence and successful uptake but only some of them showed such an effect on the rate of teacher-feedback-related comments and

learners' performance in the two tests. Some nonlinguistic cues in teacher feedback also significantly affected the occurrence and successfulness of uptake, teacher-feedback-related comments, and test results but mostly at the general level. Most of the specific types of nonlinguistic cues did not distinguish themselves from others. Metalanguage in teacher feedback had some effect on learners' noticing of the feedback and language improvement too but to a lesser degree and not always in a positive way.

Copyright by
Weiqing Wang
2009

ACKNOWLEDGEMENTS

This dissertation is impossible without the support from professors, friends, and my family. I would like to express my deepest appreciation to:

Professor Lynn Fendler, my advisor, and Professor Shawn Loewen, the co-director of this dissertation, for their careful guidance in every step I took, their detailed comments in every draft I submitted, and their continuous assurance and encouragement during difficult times.

My committee members Professor Tom Bird and Professor Charlene Polio for their insightful challenges and incisive suggestions.

My colleagues Anny Fritzen, Wenxia Wang, and Joan Zou for spending their valuable time coding part of my data for inter-rater reliability check.

The anonymous teachers and students for allowing me to observe their lessons and for making room in their tight schedule to do interviews and to take tests with me.

My friends Shih-pei Chang, Hui Jin, and Epimenio Torres Jr. for their encouraging words and generous help in the whole study.

My family for their understanding and patience throughout this long strenuous process.

TABLE OF CONTENTS

LIST OF TABLES.....ix

LIST OF IGURES.....xiv

KEY TO ACRONYMS.....xv

CHAPTER 1
INTRODUCTION.....1

CHAPTER 2
LITERATURE REVIEW AND RESEARCH QUESTIONS.....5
2.1 Introduction.....5
2.2 Mapping the big picture: input, output, and interaction.....5
2.2.1 Input.....6
2.2.2 Output.....8
2.2.3 Interaction.....11
2.3 Looking closer: feedback and the measurement of the effectiveness of feedback.....12
2.3.1 The role of feedback in second language acquisition.....12
2.3.2 The measurement of the effectiveness of feedback.....13
2.3.2.1 Tailor-made individualized tests.....14
2.3.2.2 Learner uptake.....18
2.4 Looking in particular: noticing and the measurement of noticing.....20
2.4.1 Attention and noticing in second language acquisition.....21
2.4.2 The measurement of noticing.....23
2.4.2.1 Stimulated recall.....24
2.4.2.2 Learner uptake.....28
2.5 Pulling it together: factors that may affect learners' noticing of feedback and the effectiveness of feedback.....33
2.5.1 Characteristics of feedback.....33
2.5.2 Nonlinguistic cues in feedback.....37
2.5.3 Metalanguage in feedback.....40
2.6 Summary.....42
2.7 Research questions.....44

CHAPTER 3
METHODOLOGY.....46
3.1 Introduction.....46
3.2 Teaching context.....46
3.3 Participants.....47
3.4 Instruments.....48

3.4.1 Observations....49
3.4.2 Stimulated recall interviews....50
3.4.3 Language testing....53
3.5 Procedures....57
3.6 Data analysis....60
3.6.1 The coding of observation data....60
3.6.1.1 The coding of the characteristics of feedback episodes....60
3.6.1.2 The coding of nonlinguistic cues....71
3.6.1.3 The coding of teachers' metalanguage....84
3.6.1.4 The coding of the occurrence and successfulness of uptake....86
3.6.2 The coding of stimulated recall comments....93
3.6.2.1 The coding of overall recall comments....93
3.6.2.2 The coding of language-related comments....96
3.6.2.3 The coding of target-feature-related comments....98
3.6.3 The coding of test results....100
3.6.4 The reliability of coding....101
3.6.5 Statistical analysis....102
3.7 Summary....105

CHAPTER 4
CHARACTERISTICS OF FEEDBACK EPISODES, THE NOTICING AND THE EFFECT OF TEACHER FEEDBACK....106
4.1 Introduction....106
4.2 Distribution of feedback episodes by characteristics....106
4.3 Characteristics of feedback episodes and learner uptake....108
4.4 Characteristics of feedback episodes and recall comments....119
4.5 Characteristics of feedback episodes and test results....125
4.6 Review of results....132
4.7 Discussion....134
4.8 Summary....154

CHAPTER 5
NONLINGUISTIC CUES, THE NOTICING AND THE EFFECT OF TEACHER FEEDBACK....156
5.1 Introduction....156
5.2 Occurrence and distribution of nonlinguistic cues....156
5.3 Nonlinguistic cues and learner uptake....158
5.4 Nonlinguistic cues and recall comments....165
5.5 Nonlinguistic cues and test results....169
5.6 Review of results....174
5.7 Discussion....176
5.8 Summary....192

CHAPTER 6
METALANGUAGE, THE NOTICING AND THE EFFECT OF TEACHER FEEDBACK....194
6.1 Introduction....194
6.2 Occurrence and distribution of metalanguage....194
6.3 Metalanguage and learner uptake....197
6.4 Metalanguage and recall comments....200
6.5 Metalanguage and test results....202
6.6 Review of results....204
6.7 Discussion....205
6.8 Summary....214

CHAPTER 7
CONCLUSION....215
7.1 Introduction....215
7.2 Summary and implications....215
7.2.1 Summary....215
7.2.2 Implications....216
7.3 Reflection on the present study....221
7.4 Suggestions for future research....224
7.5 Concluding remarks....225

REFERENCES....227

LIST OF TABLES

Table 1 Coding scheme for the characteristics of feedback episodes.....62

Table 2 Examples for the coding of the characteristics of feedback episodes.....70

Table 3 Coding scheme for learner uptake.....87

Table 4 Examples for the coding of the occurrence and successfulness of uptake.....90

Table 5 Examples for the coding of test results.....100

Table 6 Reliability of coding.....102

Table 7 Treatment of variables in chi-square analysis.....103

Table 8 Distribution of feedback episodes by characteristics.....107

Table 9 Overview of the occurrence and successfulness of uptake.....109

Table 10 Type and the occurrence of uptake.....110

Table 11 Type and the occurrence of uptake residuals.....110

Table 12 Type and the successfulness of uptake.....110

Table 13 Type and the successfulness of uptake residuals.....110

Table 14 Source and the occurrence of uptake.....111

Table 15 Source and the occurrence of uptake residuals.....111

Table 16 Source and the successfulness of uptake.....112

Table 17 Source and the successfulness of uptake residuals.....112

Table 18 Response and the occurrence of uptake.....113

Table 19 Response and the successfulness of uptake.....113

Table 20 Linguistic focus and the occurrence of uptake.....114

Table 21 Linguistic focus and the occurrence of uptake residuals.....114

Table 22 Linguistic focus and the successfulness of uptake.....114

Table 23 Linguistic focus and the successfulness of uptake residuals.....114

Table 24 Directness and the occurrence of uptake.....115

Table 25 Directness and the successfulness of uptake.....116

Table 26 Complexity and the occurrence of uptake.....117

Table 27 Complexity and the successfulness of uptake.....117

Table 28 Emphasis and the occurrence of uptake.....118

Table 29 Emphasis and the successfulness of uptake.....118

Table 30 Distribution of overall comments.....120

Table 31 Distribution of language-related comments.....120

Table 32 Distribution of target-feature-related comments.....120

Table 33 Type and teacher-feedback-related comments.....121

Table 34 Type and teacher-feedback-related comments residuals.....121

Table 35 Source and teacher-feedback-related comments.....122

Table 36 Response and teacher-feedback-related comments.....122

Table 37 Linguistic focus and teacher-feedback-related comments.....123

Table 38 Directness and teacher-feedback-related comments.....123

Table 39 Complexity and teacher-feedback-related comments.....124

Table 40 Emphasis and teacher-feedback-related comments.....125

Table 41 Overview of test results.....126

Table 42 Type and test results.....127

Table 43 Source and test results.....128

Table 44 Response and test results.....128

Table 45 Linguistic focus and test results.....129

Table 46 Directness and test results.....130

Table 47 Complexity and test results.....130

Table 48 Emphasis and test results.....131

Table 49 Overall results by the characteristics of feedback episodes.....132

Table 50 Occurrence of nonlinguistic cues.....157

Table 51 Distribution of different types of nonlinguistic cues.....157

Table 52 General paralinguistic cues and the occurrence of uptake.....159

Table 53 General paralinguistic cues and the successfulness of uptake.....159

Table 54 Type of paralinguistic cues and the occurrence of uptake.....160

Table 55 Type of paralinguistic cues and the successfulness of uptake.....160

Table 56 Type of paralinguistic cues and the successfulness of uptake residuals.....161

Table 57 General extralinguistic cues and the occurrence of uptake.....162

Table 58 General extralinguistic cues and the successfulness of uptake.....162

Table 59 Type of extralinguistic cues and the occurrence of uptake.....163

Table 60 Type of extralinguistic cues and the occurrence of uptake residuals.....163

Table 61 Type of extralinguistic cues and the successfulness of uptake.....163

Table 62 Type of gestures and the occurrence of uptake.....164

Table 63 Type of gestures and the successfulness of uptake.....164

Table 64 General paralinguistic cues and teacher-feedback-related comments.....166

Table 65 Type of paralinguistic cues and teacher-feedback-related comments.....167

Table 66 General extralinguistic cues and teacher-feedback-related comments.....167

Table 67 Type of extralinguistic cues and teacher-feedback-related comments.....168

Table 68 Type of gestures and teacher-feedback-related comments.....169

Table 69 General paralinguistic cues and test results.....170

Table 70 Type of paralinguistic cues and test results.....171

Table 71 General extralinguistic cues and test results.....172

Table 72 Type of extralinguistic cues and test results.....172

Table 73 Type of gestures and test results.....173

Table 74 Overall results by nonlinguistic cues.....175

Table 75 Occurrence of general metalanguage.....194

Table 76 Distribution of different types of metalanguage.....195

Table 77 Randomly selected teacher metalanguage by property and number of teachers.....195

Table 78 General metalanguage and the occurrence of uptake.....198

Table 79 General metalanguage and the successfulness of uptake.....198

Table 80 Type of metalanguage and the occurrence of uptake.....199

Table 81 Type of metalanguage and the occurrence of uptake residuals.....199

Table 82 Type of metalanguage and the successfulness of uptake.....199

Table 83 General metalanguage and teacher-feedback-related comments.....201

Table 84 Type of metalanguage and teacher-feedback-related comments..... 201

Table 85 General metalanguage and test results.....202

Table 86 Type of metalanguage and test results.....203

Table 87 Overall results by metalanguage.....204

Table 88 Uptake and teacher-feedback-related comments.....219

Table 89 Successfulness of uptake and teacher-feedback-related comments.....220

Table 90 Uptake and test results.....220

Table 91 Successfulness of uptake and test results.....220

LIST OF FIGURES

Figure 1 Main tenets of the interaction approach.....43

Figure 2 Measurement of the noticing and effect of teacher feedback.....48

Figure 3 Coding process of the observation data.....61

Figure 4 Coding process of stimulated recall comments.....94

Figure 5 Overview of the characteristics of feedback episodes predicting (successful) uptake, teacher-feedback-related comments, and correct test results.....133

Figure 6 Overview of nonlinguistic cues predicting (successful) uptake, teacher-feedback-related comments, and correct test results.....176

KEY TO ACRONYMS

ASR	Adjusted Standardized Residuals
EFL	English as a Foreign Language
ESL	English as a Second Language
L1	First Language
L2	Second Language
MA	Master of Arts
NS	Native Speaker
S	Student
SLA	Second Language Acquisition
T	Teacher
TESOL	Teaching English to Speakers of Other Languages

CHAPTER 1

INTRODUCTION

Why can't learners learn what teachers teach? What can help learners learn what teachers teach? These questions have bewildered educators and researchers for years and years and myriads of studies have been carried out to find the answer. In the area of second language acquisition, one notion that has often been used to account for the success or failure of language teaching and learning is attention. Frequently associated with the notion of attention is learners' noticing of teacher feedback and the effectiveness of teacher feedback. This is what the present study explored. Altogether, three factors were examined: the characteristics of teacher feedback episodes, the nonlinguistic cues in teacher feedback, and the metalanguage in teacher feedback. It is hoped that the results would be able to shed some light on the forever puzzling mysteries of second language teaching and learning.

There have been a plethora of studies that examined the role of the characteristics of particular feedback types (e.g, Lyster, 1998a, 1998b, 2004; Sheen, 2006) or the characteristics of whole feedback episodes (e.g., Ellis et al., 2001a, 2001b; Loewen, 2004, 2005). One strategy that has been frequently adopted by researchers to investigate the issue is to audio-tape teacher-student interactions and look at learners' verbal responses to teacher feedback as an indicator of the effect of the feedback. These studies have contributed to our understanding of factors that may influence the effect of teacher feedback. On the other hand, there is one aspect of teacher-student interactions that has

not received much attention so far. Just like in everyday conversations, oftentimes learners' responses to teacher feedback are nonverbal. In classroom settings, they might respond to what the teacher is saying with nodding, writing in their books, or using other body language. Such responses, although not as noticeable as verbal responses, constitute part of teacher-learner interactions. Will the characteristics of teacher feedback (episodes) show an effect on learners' noticing and learning when learners' nonverbal responses are taken into account? The current study attempted to fill in this gap of feedback research. I used both audio and video recordings for data collection, thus making it possible for me to identify learners' nonverbal responses to teacher feedback along with their verbal responses.

Another aspect of teacher feedback that has attracted researchers' attention is teachers' use of nonlinguistic cues. Although nonlinguistic cues have been suggested to be facilitative of language learning for almost two decades now in the L1 literature (Farrar, 1992), it was not until recently that L2 researchers started to discuss it and/or examine it in empirical studies (e.g., Davies, 2006; Faraco & Kida, 2008; Kida, 2008; Long, 2007; Sime, 2006). The results from these studies are controversial. Some researchers found that nonlinguistic cues had a positive effect on learning while others found that learners did not notice them at all. Can nonlinguistic cues have an effect on learners' noticing of teacher feedback and their language improvement after all? If yes, to what degree? These questions remain to be answered. The current study explored this area in a classroom setting. Unlike most studies which looked at nonlinguistic cues from a single angle, I looked at nonlinguistic cues from multiples perspectives, namely, learner responses to

teacher feedback, test results, and learners' retrospective comments on their perception of teacher feedback.

A third factor the current study strived to examine is the potential effect of metalinguistic terms teachers used in their interactions with students. The use of metalanguage has been under much debate for decades. To some metalanguage is not helpful to language learning and may even be counteractive (e.g., Corder, 1973) while to others metalanguage can serve as an important learning tool (e.g., Faech, 1985). Apart from such theoretical discussions, there has been very little empirical research on the role of metalanguage in language teaching and learning, especially in L2 literature. Among the few studies that have been carried out, most examined metalanguage use in terms of how teachers view it and when they use it. Little has been said about the actual effect of metalinguistic terms on learners' noticing of teacher feedback and second language learning. Is there an effect at all? Do different kinds of metalanguage have different effects? The present study sought preliminary answers to these questions, again by looking at learner responses to teacher feedback, test results, and learners' retrospective comments on their perception of teacher feedback.

Revolving around the three factors described above, the next 6 chapters of the dissertation are as follows: Chapter 2 reviews the literature on issues pertinent to the present study, specifically, a few key concepts in second language acquisition, feedback and the measurement of its effect on learning, and noticing and the measurement of noticing. The research questions are then given based on the review of literature. Chapter

3 describes methodological issues, including the settings of the study, participant information, data collection instruments and procedures, coding schemes, and statistical analysis of the data. Chapters 4, 5, and 6 respectively present and discuss the results of the study in terms of the characteristics of feedback episodes, nonlinguistic cues in teacher feedback, and teachers' use of metalanguage. Chapter 7 summarizes the results from the previous three chapters, discusses their implications, reflects on the values and limitations of the present study, and makes tentative suggestions on future research.

CHAPTER 2

LITERATURE REVIEW AND RESEARCH QUESTIONS

2.1 Introduction

This chapter reviews the literature and presents the research questions of the current study based on the review. First, three key concepts in second language acquisition are outlined: input, output, and interaction. Then, the constructs of feedback and noticing are examined in three different sections: the role of feedback and its effectiveness, the role of noticing and its measurement, and factors that may affect the noticing and effectiveness of feedback. Finally, the research questions are put forward.

2.2 Mapping the big picture: input, output, and interaction

One approach to the study of language acquisition that has been substantially discussed in SLA literature is the interaction approach (Gass, 1997, 2003; Gass & Mackey, 2006; Gass & Selinker, 2001; Long, 1996). Associated with this approach are the concepts of input, output, and interaction. The Input-Output-Interaction model has been considered to be "the closest thing that we have to a 'big' theory to date" (Block, 2003, p. 26). Gass and Mackey (2006) explain it succinctly: "the interaction approach considers exposure to language (input), production of language (output), and feedback on production (through interaction) as constructs that are important for understanding how second language learning takes place" (p. 3). Input, output, and interaction, then, are the three major constructs in the interaction approach to second language acquisition, and my study is based on insights from the interaction approach.

2.2.1 Input

Simply speaking, input refers to what is available to the learner (Corder, 1967). In the L2 literature, almost all theories of second language learning recognize the significance of input in the acquisition process (Gass & Selinker, 2001). In the behaviorist conceptualization of second language learning, input is of crucial importance because imitation is viewed as the primary mechanism of language learning. In Krashen's (1982, 1985) input hypothesis, comprehensible input, or the language information that is slightly ahead of what the language learner currently knows, is central to all language acquisition. Even in theories the focus of which is not specific input (e.g. interaction theories), input is also attributed an important role. In terms of its nature, a distinction has been made between input with positive evidence and input with negative evidence.

Positive evidence is the set of well-formed sentences the learner is exposed to while negative evidence refers to information that is provided to the learner concerning the incorrectness of an utterance (Gass, 2003). Whereas positive evidence has been unanimously regarded as a necessary condition for language acquisition, the position of negative evidence is more controversial. In the literature of first language acquisition, negative evidence has been suggested to be infrequent, often ignored, and therefore not a sine qua non for language acquisition (Baker, 1979; Brown & Hanlon, 1970; Gass & Selinker, 2001). Universal Grammar or "the system of principles, conditions, and rules that are elements or properties of all human languages" (Chomsky, 1975, p.7) has been proposed to explain the fact that first language can be acquired despite the poverty of evidence for impossible sentences. In second language acquisition, some researchers also

believe that positive evidence alone is enough for language acquisition to occur. Two most typical examples are the above-mentioned behaviorist view of second language acquisition and Krashen's (1985) input hypothesis, both of which seem to assume that acquisition will happen as long as there is the right amount and type of positive evidence. Second language acquisition, however, has been noted to be different from first language acquisition in terms of ultimate attainment, learners' prior knowledge, motivation, and so on. Because of that, some researchers have assigned negative evidence to a more important position. Schwartz (1992), for example, believes that negative evidence may affect learners' performance although it cannot permeate underlying competence. It has even been observed that adults have to be exposed to negative evidence in order to accomplish the goal of learning a second language (Birdsong, 1989; Bley-Vroman, 1989; Gass, 1988; Gass & Selinker, 2001; Schachter, 1988).

There is mounting support for the claim that negative evidence can facilitate language development (Long, 2007). One of the most widely cited study is White's (1991) investigation of how learners learn not to do something that is present in the native language but not in the target language. The participants were French learners of English who had to learn that English allows subject-adverb-verb order but does not allow subject-verb-adverb-object order, which is allowed in French. Using a pretest, immediate posttest, and delayed posttest design, White showed that negative evidence in the form of correction promoted participants' learning of adverb placement. In a follow-up study, Trahey and White (1993) provided further evidence for this conclusion. Later studies which examined the effect of corrective feedback (e.g., Ayoun, 2001; Gass & Veronis,

1989; Mackey, 2006; Pica et al., 1989; Trahey, 1996) have also lent support to White's findings. Some of these studies even showed that negative evidence can be more effective in certain cases. A typical example is Iwashita's (2003) investigation of the differential effects of negative feedback and positive evidence on beginning Japanese as a second language learners' short-term grammatical development. The data were drawn from the task-based interactions of 55 native-nonnative speaker dyads. Five types of interactional moves were categorized either as negative feedback (recasts and negotiation) or as positive evidence (completion, translation, and simple model). It was found that although positive evidence occurred 10 times more than negative feedback, only learners who had an above-average score on the pretest benefited from it. Implicit negative feedback, by contrast, had beneficial effects on learners' short-term grammatical development irrespective of their current mastery level of the targeted language structures.

With all that said, the role of positive and negative evidence in second language acquisition can be summarized in the following way:

> positive evidence can reveal to learners the presence of information in the second language that is different from the native language, but...negative evidence is necessary to show what is not possible in the second language when it is possible in the native language. (Gass & Selinker, 2001, p. 284)

2.2.2 Output

Output is the language produced by the learner (Gass & Mackey, 2006). It has been claimed to be another condition for successful language learning. Output is regarded as a way to practice new knowledge that has been learned. Another role it has been assigned is that it provides a forum for learners to receive more input by eliciting feedback. The

concept of output did not receive heavy attention until Swain's (1985) proposal of the output hypothesis. This hypothesis is a response to the fact that students in Canadian immersion programs demonstrated weaknesses in grammatical accuracy even though they received rich comprehensible input. It holds that if students are to be both fluent and accurate in the target language, what they need is not only comprehensible input but also opportunities to get feedback on what they produce. A key concept in the hypothesis is *comprehensible* output or *pushed* output. What is meant by this notion is that learners are "forced" to produce more accurate language as a necessary part of making themselves understood. Swain explains it this way: when learners experience comprehension problems, they will be pushed to make their output more precise, coherent, and accurate, and thus be forced "to move from semantic processing to syntactic processing" (p. 249). In a more recent article, Swain (2005) specifies three functions of learner output: the noticing function, the hypothesis-testing function, and the metalinguistic function. The design of my study is based on these understandings of input and output.

While producing language, learners may notice that they do not know how to say what they want to say or that they only have partial knowledge to express themselves. If this gap is (made) explicit enough, it will raise learners' awareness of their language problems. For example,

(1) S: I went to a...a zoo.
T: What zoo?
S: A zoo with lots of sea animals and fish.
T: An aquarium?
S: Oh, you call that an aquarium?

In this scenario, the student's response "you call that an aquarium?" is a good sign of selective attention and learning. It is through her language production and the feedback it elicits that she notices her lack of the vocabulary word "aquarium."

Sometimes learners are not sure of certain language structures. By producing language, they can try out their hypotheses about such structures. With the feedback from interlocutors confirming or disconfirming their hypotheses, they may arrive at a conclusion. In the following example, a Chinese ESL learner is not sure of the English word "eye-opening," the literal translation of which into Chinese is "opening-eye," so she tries it out:

(2) S: This is a...opening-eye experience for me.
T: You mean an eye-opening experience?
S: Oh, eye-opening, not opening-eye?

Through this short conversation, the learner might realize that literal translation is not going to work here. This realization is achieved by her trying her hypothesis about the word "opening-eye" and the negative-evidence feedback the teacher gives her.

According to Swain, output also enables learners to control and internalize linguistic knowledge and causes them to reflect on their own use of the target language, which is a possible step toward language acquisition. For example:

(3) S: I saw two deers in the woods.
T: You saw two deer?
S: Oh, deer is an irregular plural form?

From her meta talk, the learner is obviously thinking metalinguistically about the plural form of the word "deer". The difference between her utterance and that of the teacher has

triggered this reflection. As a result, she has a better chance to learn the correct plural form of "deer". Without her producing "deers", this might not have happened.

Since the proposal of the output hypothesis, various studies (e.g., Ano, 1998; McDonough, 2005; Swain, 1997; Swain & Lapkin, 1995) have lent support to it. Discussing how output provides a forum for language learning, Gass and Selinker (2001) list four functions of output: (1) testing hypothesis, (2) receiving crucial feedback for the verification of these hypotheses, (3) developing automaticity in interlanguage production, and (4) forcing a shift from more meaning-based processing of the second language to a more syntactic mode. These functions, although expressed in a different way, are in keeping with Swain's output hypothesis.

2.2.3 Interaction

Decades ago, Wagner-Gough and Hatch (1975) argued that conversational interaction is the basis for the development of syntax, not just a forum for the practice of grammatical structures. It is now commonly accepted in the field of SLA that interaction and learning are closely related.

Conversational interaction can be defined as the exchanges between/among learners or between/among learners and native speakers where negotiation of meanings occurs (Gass & Mackey, 2006). Central to this definition is the notion of negotiation of meaning, which refers to instances in which participants interrupt the flow of the conversation in order for both parties to understand what the conversation is about (Gass & Selinker,

2001). The role of negotiation of meaning in interaction is illustrated in Long's (1996) frequently cited statement:

> Negotiation of meaning, and especially negotiation work that triggers interactional adjustments by the NS or more competent interlocutor, facilitates acquisition because it connects input, internal learner capacities, particularly selective attention, and output in productive ways. (p. 414)

From this statement, it can be seen that negotiation of meaning involves both input and output, two of the most important factors in language learning as discussed above. Given that, it is not surprising that Long's (1983, 1985, 1996) interaction hypothesis has been said to subsume aspects of both Krashen's (1982, 1985) input hypothesis and Swain's (1985, 1995, 2005) output hypothesis (Gass & Mackey, 2006). Gass (1997) has also argued for the facilitating role of interaction through negotiation of meaning. In her view, focused negotiation work in interaction can orient learners' attentional resources to what they do not know or do not know well, thus serving as a priming device to set up the initial step of learning.

2.3 Looking closer: feedback and the measurement of the effectiveness of feedback

The role of interaction in SLA has invited a large body of research (e.g., Ellis et al., 1994; Gass & Veronis, 1994; Polio & Gass, 2005). A concept that has received heavy attention in interaction research is feedback. This includes both the effect of feedback on learners' language improvement and the measurement of the effect.

2.3.1 The role of feedback in second language acquisition

Feedback in conversational interactions can provide input to learners either as positive evidence or negative evidence or a combination of the two, depending on the nature of

the feedback. Upon receiving feedback, learners may try to correct themselves. The new utterance can be regarded as pushed output, which can then help language learning by realizing the three functions illustrated by Swain (2005). Long (2007), in a new book about problems in SLA, devotes a whole chapter to the role of recasts, a type of implicit negative feedback. After a detailed review of both cross-sectional and longitudinal studies, Long concludes that, although they are not necessary for acquisition, recasts "appear to be facilitative, to work better than models, and to do so incidentally, without interrupting the flow of conversations and participants' focus on message contents" (p. 94). This facilitative role of feedback in language learning is also reflected in negotiation:

> Negotiation serves as a catalyst for change because of its focus on incorrect forms. By providing learners with information about incorrect forms, negotiation enables learners to search for additional confirmatory or nonconfirmatory evidence. (Gass & Selinker, 2001, p. 283)

While the focus of this claim is negotiation, it points to the importance of feedback. To a large extent, feedback is what initiates changes in learners' interlanguage systems. Without feedback, learners may not notice the gaps or holes in their L2 knowledge, may not modify their language production, and may not get the input they need in order to confirm or disconfirm their hypotheses about a language structure. Both the aforementioned studies that examined the role of negative evidence and those reviewed later about the effectiveness of feedback (see Section 2.5) strongly suggest that feedback plays a significant role in second language learning.

2.3.2 The measurement of the effectiveness of feedback

Loewen (2004) provides a list of the methods that have been utilized to measure the effectiveness of recasts (and other types of feedback) in classroom contexts: learner

uptake, uptake charts following a lesson, spontaneous production of recast forms elicited in task-based interactions, elicited production of targeted forms in tailor-made individualized tests, and indirect measures such as grammaticality judgment tests that seek to access learners' underlying knowledge of a form. All these methods have their own advantages and disadvantages. For the purpose of the current study, only two methods are reviewed: tailor-made individualized tests and learner uptake.

2.3.2.1 Tailor-made individualized tests

Tailor-made individualized tests have been discussed and adopted by a number of researchers who study interaction (e.g., Loewen, 2002, 2005; Loewen & Philp, 2006; Nabei & Swain, 2002; Swain & Lapkin, 1998; Williams, 2001). Since they are used in different contexts by different researchers, no universal rules have been established about how to develop such tests and what forms they should take. Overall, tailor-made individualized tests tend to require learners to correct their nontarget-like utterances on which they have previously received feedback (Loewen & Philp, 2006). Test items are created from forms that arise in language episodes which involve test takers. Below is an example which shows how tailor-made test items are developed.

(4)

Language episode:

OL: He want to hitchhike…travel and–
L: travel to Europe…hitchhiking? What hiking?
OL: Hitchhiking, you know?
L: no.
OL: Like you don't have car but I want to move another city and my thumb –
L: – Oh! OK hitch. How do you spe-no – how do you pronounce it?
OL: Hitchhike.
L: Hitchhike.

Test item:

Lena wanted to travel to the next town but she had no car and no money for the

bus. So she decided to _____. She walked to the side of road and put out her thumb. Soon a car stopped and gave her a ride. (From Williams, 2001, p. 331)

In the language episode, a learner initiates a question about a vocabulary word and another learner supplies the information. The test item accordingly assesses whether the initiator of the question can recall the word by filling in a blank.

Like any language tests, tailor-made individualized tests are not perfect. One of the biggest disadvantages of such tests is that since test items are based on language forms that incidentally occur in language episodes, they cannot be predicted or are hard to predict. As a result, it is difficult to design pretests. In the absence of pretests, however, the tests cannot provide information about learners' previous knowledge of the forms tested and thus cannot differentiate between the acquisition of new knowledge and the consolidation of latent knowledge. As Loewen (2002) argues, for individualized testing, "the FFE [focus-on-form episode] itself has to serve as a type of pretest, indicating students' lack of prior knowledge or ability regarding the targeted linguistic item" (p. 83). On the other hand, since they are specific to particular language forms and particular learners receiving particular feedback, such tests can at least "inform us about learners' ability to use demonstrably problematic forms subsequent to the provision of feedback" (Loewen & Philp, 2006, p. 542). Individualized testing therefore can serve as a great tool to measure the effectiveness of feedback.

One issue that is pertinent to any type of language testing is construct validity. In the area of SLA, however, there is a widespread failure to consider the construct validity of testing instruments (Ellis, 2005). As a relatively new testing method, and because of their

complexity and variation in terms of their contents, format, etc., tailor-made individualized tests have received even less attention with regard to their construct validity. One study that deserves special attention in this respect is Loewen's (2002) study of incidental focus on form. In order to design a valid and reliable testing instrument, Loewen conducted two pilot studies and performed a validity and reliability analysis on the individualized test items in his final study. Despite such careful design and analysis, Loewen still showed a concern with the construct validity of his tests, saying that "Another limitation related to the testing concerns whether learner performance on the test items are representative of learner performance in other contexts" (p. 287). When "other contexts" are classroom interactions, Loewen's concern can be interpreted as the mismatch between learner performance in the tests and learner performance in the classroom. Specifically, the tests most likely allowed learners to access their explicit L2 knowledge. However, this knowledge could be much less accessible during spontaneous classroom interactions due to the different settings and objectives of learner language production. This discrepancy between test performance and classroom performance points to two important issues concerning tailor-made individualized tests: what to test and how to test. Both are crucial questions in examining the construct validity of such tests. The what question can virtually be traced to the relationship between implicit knowledge and explicit knowledge while the how question is a matter of how to operationalize the two types of knowledge for research purpose.

Ellis (2005) illustrates the relationship between implicit knowledge and explicit knowledge from three different interface theories. The noninterface position largely rejects the possibility of explicit knowledge transforming directly into implicit knowledge

and vice versa. The strong interface position claims that the two types of knowledge can be derived from or converted into each other. The weak interface position acknowledges the possibility of explicit knowledge becoming implicit but with some limitation on when or how this can take place. Despite the fact that there are different and conflicting positions about the role played by explicit knowledge in the acquisition of implicit knowledge, it is widely accepted that explicit knowledge can contribute to performance. A test that supposedly measures learners' implicit knowledge, therefore, may in actuality measure some aspects of learners' explicit knowledge. Meanwhile, it has been found that different performance tasks are likely to induce L2 learners to draw differentially on their implicit and explicit knowledge (Bialystok, 1982, cited in Ellis, 2005). This means that the two types of knowledge can probably be measured in different ways.

Based on the characteristics of implicit knowledge and explicit knowledge, Ellis (2005) proposes seven criteria for assessing L2 implicit knowledge and explicit knowledge: degree of awareness, time available, systematicity, certainty, focus of attention, metalinguistic knowledge, and learnability. More specifically, when a task assesses learners' implicit knowledge, learners make use of "feel" when responding to the task; they are pressured to perform the task online; they provide consistent answers; and they are highly certain of their responses; the primary focus of the task is meaning; it does not require learners to use metalinguistic knowledge, and the knowledge involved favors early learning. In contrast, when a task assesses learners' explicit knowledge, learners provide responses to the task using rules; they are not pressured by time when performing the task; they provide variable responses; and they have a low degree of certainty in their

responses; the primary focus of the task is form; it encourages learners to use metalinguistc knowledge; and the knowledge involved favors late, form-focused instruction. In a psychometric study which aimed to develop a battery of tests that would provide relatively separate measures of implicit and explicit knowledge, Ellis (2005) found that most of these criteria proved true although some of them did not receive clear evidence or were not supported because of various factors such as learners' random behavior.

Going back to Loewen's (2002) concern about his individualized tests, it suggests that the items in such tests should tap into learners' implicit knowledge (e.g, knowledge accessed during classroom interactions and online production), not just their explicit knowledge (e.g, knowledge accessed in decontextualized tests). This requires investigators to minimize the mismatch between the testing situation and classroom interactions.

2.3.2.2 Learner uptake

As early as the 1980s, the term "uptake" was proposed by Allwright (1984) to refer to what learners report they have learned from a lesson. More recently, Lyster and Ranta (1997) define uptake as "a student's utterance that immediately follows the teacher's feedback and that constitutes a reaction in some way to the teacher's intention to draw attention to some aspect of the student's initial utterance" (p. 49). Uptake is regarded as an important observable source for understanding the impact of feedback (Nabei & Swain, 2002), hence it has been used to measure the effectiveness of feedback in various studies (e.g., Ellis et al., 2001a, 2001b; Loewen, 2002, 2004; Lyster, 1998a; Lyster &

Ranta, 1997; Panova & Lyster, 2002; Sheen, 2006. For more detail about these studies, see Section 2.5). According to Ellis et al. (2001a), as also discussed in Basturkmen et al. (2002) and Loewen (2004), "there are theoretical grounds for believing that uptake might contribute to acquisition" (p. 287). Two reasons are given for this claim. First, uptake as a response to teacher feedback helps learners practice language items and automatize the retrieval of them. Second and more importantly, by producing uptake, learners attempt to use forms they have previously used incorrectly or received information about. Uptake in this sense is a type of pushed output, which allows learners to process syntactically rather than semantically. Uptake therefore "may create the conditions needed for language acquisition to occur" (Ellis et al., 2001a, p. 287).

On the other hand, some researchers (e.g., Long, 2007; Ohta, 2000) argue that uptake is merely a discourse phenomenon that may not necessarily be related to the acquisition process. Uptake is an optional move (Ellis et al., 2001a; Loewen, 2004; Sheen, 2006). Learners may or may not respond to the linguistic information provided to them. At times they may choose not to produce uptake even if there is an opportunity for them to do so. This may happen when other interactional tasks are in priority or when they think it is unnecessary to respond. There are also occasions where learners do not have the chance to respond to the feedback they receive. Many studies (e.g., Loewen, 2002, 2004; Lyster, 1998a; Oliver, 2000) have shown that teachers may continue their turn after giving learners feedback, allowing no time for learners to respond to the feedback. Uptake is therefore not a guarantee of acquisition (Williams, 2001). Even researchers who have repeatedly used uptake in their studies have been very cautious with the term. Ellis et al.

(2001a), for example, reiterate that "uptake cannot be viewed as evidence that acquisition has taken place...nor...necessary for acquisition to take place" (p. 287). In view of such concerns, uptake alone may not be sufficient to examine the effectiveness of feedback and may work better together with other measures.

2.4 Looking in particular: noticing and the measurement of noticing

The important role of feedback in second language learning is evident. The degree of the effectiveness of feedback, however, is more complicated. To a large extent it depends on learners' noticing of feedback and yet it is not always clear whether learners can perceive feedback as expected (Gass & Selinker, 2001; Hawkins, 1985). Below is an example where the learner fails to recognize a clarification request as negative evidence.

(5) S: One day, the fairy, sting the magic wand to Cinderalla.
T: Sorry?
S: One day, the fairy sting the magic wand to Cinderalla.
T: Ok.
S: Cinde, ah, Cinderaella changed into, the beautiful girl. (Laugh) Ah, and, the, Cin, Cinderella went to the palace by coach. The, the prince fall in love at a first glance.
T: Sorry?
S: Ah, the prince fall in, falled falled in love Cinderella at a first glance. And they dance, they danced... Ah, Cinderella have, Cinderella have to go home. (Takashima, 1995, cited in Gass and Mackey, 2006)

In this example, although the learner might have finally been alerted to the use of past tense as indicated in his struggle to change "fall" into "falled" and "dance" into "danced", he does not make any change after the first "sorry". He probably has assumed that the teacher does not understand him simply because of his low voice, unclear pronunciation, or broken discourse.

Given the complexity of noticing, this section first briefly talks about the role of attention and noticing in second language acquisition, and then discusses the measurement of noticing.

2.4.1 Attention and noticing in second language acquisition

Attention has been identified as an important cognitive process in second language acquisition, a process that "encodes language input, keeps it active in working and short-term memory, and retrieves it from long-term memory" (Robinson, 2003, p. 631). Noticing is the part of the attentional system that involves the detection and consequent registration of stimuli in memory (Philp, 2003; Robinson, 1995; Tomlin & Villa, 1994). The role of attention and noticing in selecting input for L2 learning is a contentious issue in the SLA literature. Some researchers believe that attention is not important in second language acquisition. A typical example is Krashen. Krashen (1982) made a distinction between language acquisition, a subconscious process, and language learning, a conscious process. He believes that only the acquired system can be used to produce language while the learned system can only serve as a monitor for the former. Unlike Krashen, some researchers have seen an important role of attention and noticing in SLA. Long (1996), for example, argues that selective attention along with learners' L2 processing capacity mediates the L2 acquisition process. VanPatten (1994) demonstrated the role of attention in L2 acquisition by showing how a Spanish learner starts to use subjunctive mood in his own speech after he notices it in others' speech. Schmidt (2001), one of the strongest proponents for the role of attention and noticing, believes that "the concept of attention is necessary in order to understand virtually every aspect of second

language acquisition" (p. 3). In his famous noticing hypothesis (1990, 1993, 2001), it is proposed that the subjective experience of "noticing" is the necessary (and sufficient) condition for the conversion of input to intake; noticing is therefore the first step of language learning. A series of studies have been conducted around the issue of attention and noticing (e.g., Kim, 1995; Leow, 1997, 2000; Robinson, 1996, 1997; Rosa & O'Neill, 1999; Schmidt & Frota, 1986). Although some studies lend less support to the noticing hypothesis, the cumulative findings show that attention and noticing *are* important in SLA. It has been argued that even if noticing is not necessary for L2 learning, it is facilitative. Even Gass (1997), who attributes some learning to universal grammar and argues against the claim that all L2 learning requires attention, stresses that her argument is not intended to lessen the importance of attention. In simple words, "there does not appear to be any evidence at all against the...claim that people learn about the things they attend to and do not learn much about the things they do not attend to" (Logan et al., 1996, cited in Schmidt, 2001, p. 30).

In the SLA literature, the term "noticing" has been associated with and distinguished from other concepts such as awareness, alertness, consciousness, and detection. Robinson (1995), for example, distinguishes between noticing and detection on the basis of awareness, defining noticing as detection with awareness and rehearsal in short-term memory. For the purpose of the current study, noticing refers to detection accompanied by lesser and greater degrees of awareness (Philp, 2003). The various levels of awareness; however, were not explored respectively.

2.4.2 The measurement of noticing

In the SLA literature, noticing has been measured through both online procedures and offline procedures. Online procedures or concurrent reports take various forms such as think-aloud protocols (e.g., Alanen, 1995; Loew, 1997, 2000), private speech (e.g., Ohta, 2000), online uptake charts (e.g., Mackey et al., 2001), and online learning journals (e.g., Mackey, 2004, cited in Egi, 2004). Offline procedures or retrospective reports also take a variety of forms such as post-treatment questionnaires (e.g., Mackey et al., 2002; Robinson, 1995), stimulated recall (e.g., Egi, 2007a, 2007b; Mackey et al., 2000; Swain & Lapkin, 1998, 2001), offline uptake charts (e.g., Slimani, 1989), prompted repetition of feedback (e.g., Philp, 2003), and diaries (e.g., Schmidt & Frota, 1986). When investigating learners' noticing of feedback and its impact on learning, it is important to make sure that the measures of noticing "should accurately capture learners' cognitive processes while neither facilitating nor hindering learning" (Egi, 2004, p. 243). As researchers strive for this goal, concerns have been raised about both online and offline procedures (e.g., Mackey, 2006; Schmidt 2001). For example, online think-aloud protocols require learners to report their mental processes under temporal and communicative pressure. This can lead to the underreporting of noticing and thus fail to capture certain aspects of the cognitive processes. Offline procedures such as diaries and questionnaires have their own problems too. Tomlin and Villa (1994) point out that the cognitive processing of input "takes place in relatively brief spans of time, seconds or even parts of seconds" (p. 185). However, diaries and questionnaires might span a much longer time. By the time learners write their diary or do the questionnaire, they might have forgotten what they were thinking. Such reports of noticing, like online think-aloud

protocols, may also lead to underreporting. In view of such problems, Mackey (2006) and Mackey and Gass (2006) suggest that researchers should triangulate methods of collecting noticing data to obtain a picture as full as possible.

The advantages and drawbacks of verbal reports have been substantially discussed by Egi (2004, 2007a, 2008). Two major issues concerning the validity of verbal reports in Egi's discussion are reactivity and veridicality. Reactivity concerns the effect of verbalization on participants' task performance. If verbal reports affect participants' task performance, they are reactive. If not, they are non-reactive. Veridicality concerns the accuracy of verbal protocols as a reflection of learners' cognitive processes. If verbal reports fail to capture participants' thoughts or include thoughts that did not take place, they are non-veridical. In line with the principle that verbal reports should not influence participants' performance while accurately reflecting their cognitive processes, "non-reactive and veridical reports are considered to be valid data" (Egi, 2004, p. 245). For the purpose of the current study, two noticing measures are discussed in view of their reactivity and veridicality: stimulated recall and again learner uptake.

2.4.2.1 Stimulated recall

Stimulated recall is a retrospective data collection method in response to the ongoing shift toward examining cognitive factors in second language research (Gass & Mackey, 2007). It is often conducted by giving participants stimuli and asking them to comment on their thoughts during a previous task. Such comments allow researchers to gain access to participants' internal thinking processes and gather information that might be

unavailable from observation data. Like other retrospective verbal reports, it has been examined with respect to reactivity and veridicality.

Reactivity is not an issue for stimulated recall when it is administered after a task (Egi, 2004). When it is carried out before a task, however, it can be reactive due to at least two reasons: stimulus presentation and verbalization (Egi, 2008). While recall cues such as video or audio recordings of participants' task performance can help them more completely recall their cognitive processes and thus mitigate the problems caused by memory decay, they can also serve as additional L2 input. As a result, learners are exposed to the same L2 input twice, once during the task and once during the recall. This double-input exposure (Leow, 2002) may reinforce learners' previous learning or give them time to process input that they have not previously processed. The second source of reactivity is the act of verbalization. Egi (2004) explains it with Swain's (1985, 1995, 2005) output hypothesis. For one thing, verbalization of the cognitive process gives learners additional opportunities to practice the target language, which can promote fluency. For another thing, verbalization gives learners opportunities to reflect on their own language use, thus fulfilling the metalinguistic function of pushed output. For such reasons, both the stimuli and verbalization can affect learners' test performance. In terms of veridicality, there are also two potential sources for the non-veridicality of stimulated recall. The first source again is stimulus presentation. As described above, the recall cues provide learners with additional input. This double exposure to the same input may allow learners to reconstruct their thoughts at the time of recall rather than retrieving information from long-term memory (Egi, 2004; Gass & Mackey, 2000). The second

source is what Egi (2004) calls the interviewer effect. Stimulated recall is often carried out in an interview format. The presence of the interviewer may lead participants to report what they believe is of interest to the interviewer and/or avoid reporting their errors in order to sound intelligent (e.g. Jourdenais, 2001; Norris,1990, cited in Egi, 2004). In brief, stimulated recall faces reactivity and veridicality issues depending on when and how it is conducted.

In the area of SLA, although stimulated recall has been adopted by many researchers to measure cognitive processes such as noticing and awareness, its validity is under-explored (Egi, 2004, 2008). One study that has shed light on the reactivity of stimulated recall is the one by Adams (2003). Fifty-six L2 Spanish learners collaboratively wrote a story in pairs. They were then randomly assigned to three groups: a control group, a noticing-only group, and a noticing-plus-stimulated-recall group. The control group repeated the writing task without additional treatment. The noticing-only group compared their original essay and a reformulated version of their essay. In addition to text comparison, the noticing-plus-stimulated-recall group recalled their thoughts at the time they made the text comparisons. Both experimental groups then individually wrote the story again. Results indicated that the noticing-plus-stimulated-recall group incorporated reformulations in the second writing significantly more than the noticing-only group. It was concluded that learners' participation in the stimulated recall had benefited their learning, suggesting that stimulated recall in this study was reactive. Egi's (2007a) study provided support for the "contaminating effect" of stimulated recall. Two groups of learners were asked to report their interpretation of recasts provided during

conversational interactions: an immediate report group and a stimulated recall group. The immediate report group reported any thoughts they had during a brief conversational turn as prompted by a native speaker interlocutor (for details about immediate reports, see Egi, 2004, 2007a and Philp, 2003). The stimulated recall group conducted a recall interview after the interaction sessions while watching their own videotaped interactions and communicative tasks. The two groups did not show significant differences on the immediate posttest, which was administered before the stimulated recall session. However, in the delayed posttest, which was administered after the stimulated recall session, the stimulated recall group significantly outperformed the immediate report group. This change suggests that the stimulated recall interview had influenced learners' performance. Like Adams' study, this study shows that stimulated recall can be reactive if it is conducted before a testing event.

Compared with its reactivity, the veridicality of stimulated recall has received less attention, probably because one of the major criticisms of delayed retrospective reports, memory decay, is alleviated by stimulus presentation. To look into this issue, we can turn to another piece of research by Egi (2004). Although the main purpose of the study is to explore the use of immediate retrospective verbal reports, it also has indications for the veridicality of stimulated recall. The design of the study is similar to but different from the one described above. It followed a pretest–treatment–posttest sequence, but there was no delayed posttest. Participants were also divided into two groups: an immediate report group and a stimulated recall group. Similarly, the immediate report group recalled their thoughts about language episodes immediately after a brief conversational turn as

prompted by an auditory stimulus while the stimulated recall group recalled their thoughts while watching videotaped treatment sessions after the completion of the treatment and posttest. Analysis of the data suggested that stimulated recall was less effective in capturing learners' noticing due to various factors such as memory decay, multiple exposures to the same input, and interviewer effect. This result suggests that stimulated recall is not always veridical.

In summary, although stimulated recall can be argued to elicit valid data when carefully operationalized (Egi, 2004), there are problems researchers need to be careful with whether it is conducted before or after testing events. Measures should be taken to reduce or circumvent its reactivity and non-veridicality.

2.4.2.2 Learner uptake

Uptake in its traditional definition as verbal report is not free from reactivity and non-veridicality. For example, in the use of online uptake charts, learners' normal learning activities might be interrupted, which may in turn affect their test performance. In the use of offline uptake sheets, there might be new information that learners have noticed but are not able to recall at the moment of reporting. In either case, learners may report what they have not actually comprehended in order to appear intelligent. Such reports can provide researchers with false information about noticing and learning.

In its more recent definition by Lyster and Ranta (1997, see Section 2.3.2.2), uptake is a learner's response to the teacher's corrective move, the final objective of which is the

learner's attention to his or her erroneous utterance. The production of uptake, successful uptake in particular, may therefore indicate that the learner has indeed attended to his or her erroneous utterance as expected. In the SLA literature, many researchers have referred to uptake as an indication of noticing. Lightbown (1998), for example, claims that "a reformulated utterance from the learner gives some reasons to believe that the mismatch between learner utterance and target utterance has been noticed, a step at least toward acquisition" (p. 193). Believing that uptake constitutes one type of pushed output, Loewen (2004) observed that "This pushed output, then, may be an indication of noticing" (p. 157). Loewen and Philp (2006), talking about the relationship between uptake and the noticing of recasts, stated that "the production of successful uptake provides an indication that the learner has noticed the recast" (p. 542). Sheen (2006), drawing on previous discussions about uptake as an indication of noticing and studies that probed into the relationship between uptake and the noticing or interpretation of recasts, concluded that "uptake/repair can serve as one measure of learner noticing and thus have potential for language acquisition" (p. 368).

To me, uptake has methodological advantages that other noticing measures may not have. First, unlike think-aloud protocols, uptake is part of interactional discourse and does not require learners to stop the task at hand to report what is going on in their mind. The occurrence of uptake therefore will not interrupt the learning process and affect ultimate test performance. For the same reason, the reactive effect posed by the verbalization of cognitive processes can be somewhat circumvented. Furthermore, uptake does not involve stimulus presentation. This helps to solve the double-input exposure problem

caused by recall cues. In terms of veridicality, uptake has its strengths too. Uptake occurs online. As a result, memory loss is not a problem. Since no interviewer is involved, learners do not need to provide information that they think is of interest to the interviewer but that does not actually reflect their real thinking. Interviewer effect therefore is not an issue either. In brief, although uptake may not provide information that directly describes the cognitive process, it can at least give researchers clues about whether noticing has happened and sometimes, what has been noticed.

Apart from theoretical discussions, empirical evidence for the potential of uptake as a measure of noticing can be found in both L1 and L2 research. In first language acquisition, it has been observed that children frequently repeat recasts (See Long & Robinson, 1998). Repetition constitutes a type of uptake. Although learners may repeat teachers' model utterances as nothing more than a rote language behavior (Mackey & Oliver, 2002), it should not be ruled out that sometimes repetition does result from learners' noticing of the corrective nature of recasts and the errors in their initial utterances. In second language acquisition, one noteworthy study pertinent to uptake as a noticing measure is Mackey et al's (2000) investigation of learners' recognition of the target of feedback. The participants first finished a spot-the-difference-task and then watched the videotape of their performance and made comments on it. In a post hoc analysis, the researchers explored the relationship between learners' perception about feedback and their immediate uptake of the feedback during the interaction. It was found that learners' stimulated recall reports generally revealed accurate perception about feedback for which they had uptake at the time of the interaction. Specifically, in cases

where learners produced uptake, they also seemed to be accurate in their perception of what the feedback was about. When they did not react to the feedback in the discourse, they generally did not perceive it as feedback. This finding provides strong support for uptake as an indication of noticing.

Han (2002) drew on learner uptake of feedback to look at noticing. In her small-scale empirical study about the impact of recasts on tense consistency in L2 output, Han gave an example about how a learner's tense consistency in an oral narrative improved after she received recasts. Han discussed this change with reference to uptake and noticing.

> In producing the oral narrative above, Jee-Young received six recasts, four of which targeted her incorrect use of tense. Her uptake suggests that her attention was indeed drawn, albeit to a varying extent, to the discrepancies between her own output and the input provided by the recasts. (p. 561)

Here Han's conclusion about the learner's noticing of the differences between the tense use in her own production and that in the recasts was drawn from the learner's uptake. Although uptake was not clearly stated as the measuring tool of noticing in the study, it can be seen that Han was using it to make inferences about noticing.

Han (2002) is not the only one who interpreted uptake as evidence for noticing. In a study about the effectiveness of recasts in different instructional settings, Lyster and Mori (2006) concluded that "the overall higher proportion of uptake and repair following feedback in JI [Japanese immersion] classrooms ...suggests that JI students were predisposed to noticing the corrective purpose of recasts" (p. 292). This statement suggests that Lyster and Mori also treated uptake as an indication of noticing in their study. A study which took learner uptake as the primary measure of noticing is Song

(2007), who argued that uptake "may be an effective way of assessing noticing" (p. 6) given its role as pushed output. Operationalizing noticing with uptake, Song found that learners' noticing of recasts was constrained not only by the number of changes in recasts, but also by the linguistic domains of those changes.

Despite the belief in the potential of uptake as an indication of noticing, the use of uptake as a measurement for noticing is a contentious issue for the same reasons as described in Section 2.3.2.2: its optional nature and sometimes the lack of opportunities for its production. As a result, while successful uptake can provide evidence for noticing, the reverse does not necessarily hold true (Sheen, 2006). Furthermore, one may argue that since uptake is an internal part of interactional discourse and a type of pushed output, it inevitably has an effect on testing results and hence is inevitably reactive.

In summary, both stimulated recall and uptake have advantages and disadvantages as measures of noticing. Stimulated recall may directly provide specific information about learners' internal thought processes, but may be reactive and non-veridical due to factors such as stimulus presentation, verbalization, memory decay, and interviewer effect. Uptake may circumvent validity problems caused by these factors, but may not provide complete and direct information about cognitive processes due to its optional nature and sometimes the lack of opportunities for learners to produce it. Taken separately, neither stimulated recall nor uptake can be assumed to measure noticing with complete accuracy. Taken together, however, they can be supplementary to each other. Uptake may provide information that learners have forgotten even with recall cues while stimulated recall may

provide information that is not observable from uptake. In the hope to get a fuller picture of learners' noticing of teacher feedback, and also to answer the call for method triangulation in collecting noticing data, both stimulated recall and learner uptake were adopted in the present study.

2.5 Pulling it together: factors that may affect learners' noticing of feedback and the effectiveness of feedback

Having discussed the role of noticing and feedback in second language acquisition, this chapter now turns to factors that might affect learners' noticing of feedback and the effectiveness of feedback. Among the various factors, three are discussed: the characteristics of feedback, the nonlinguistic cues in feedback, and the metalanguage in feedback. It should be noted that in most of the studies mentioned below, uptake was used as a major indicator for the effectiveness of feedback, not a noticing measure.

2.5.1 Characteristics of feedback

The effectiveness of feedback has been found to be related to the specific features of feedback. A study by Loewen and Philp (2006) revealed that recasts with stress, recasts with declarative intonation, recasts in extended episodes and recasts with only one change led to more successful learner uptake and higher test scores. Sheen (2006) examined seven characteristics of single-move recasts: mode, scope, length, reduction, number of change, type of change, and linguistic focus. It was found that recasts that were short, declarative, reduced, repeated, with a single error focus, and involving substitutions were positively related to learner uptake and repair.

These studies were focused on the characteristics of a particular type of feedback, namely recasts. With the rise of focus on form research, there have been a series of studies which investigated the effect of the characteristics of focus-on-form episodes (e.g., Ellis et al., 2001a; 2001b; Loewen, 2003, 2004, 2005). A focus-on-form episode often consists of an occasional shift of attention to linguistic code features (Long & Robinson, 1998). It is triggered by comprehension or production problems, whether they are perceived or anticipated (for a detailed discussion, see Williams, 2005). As a result, focus on form often engages different types of teacher feedback. Studies on the characteristics of focus on form, therefore, shed light on the characteristics of various feedback types. Two of the most important studies are Ellis et al. (2001a) and Loewen (2004). In Ellis et al., four characteristics of focus-on-form episodes were investigated: source, complexity, directness, and linguistic focus. Results indicated that there was a higher rate of uptake in negotiation of meaning than form and that it occurred more often and was more successful in complex than simple episodes. Directness and linguistic focus, on the other hand, were not significant factors although uptake was more successful in episodes involving pronunciation than those involving vocabulary. Using a similar method, Loewen (2004) investigated the occurrence of uptake and its successfulness in relation to the characteristics of incidental focus on form. He found that complex, immediate, and elicit feedback led to a high level of uptake, and the successfulness of uptake was associated with complex, code-related, immediate, reactive, heavy, and elicit feedback.

All the studies reviewed above demonstrate that the characteristics of feedback have an impact on the effectiveness of feedback. In view of the important role of noticing in second language learning, the characteristics of feedback may also be related to noticing. In actuality, some researchers have investigated this relationship. Following his finding that recasts led to a low level of uptake, Lyster (1998b) explored the reason why it was as such. Two characteristics of recasts were examined: whether they were isolated or incorporated, and whether they were declarative or interrogative. Recasts were coded into four categories: isolated and declarative, isolated and interrogative, incorporated and declarative, and incorporated and interrogative. Lyster found that recasts were distributed in equal proportions and fulfilled identical functions with noncorrective repetitions and that teachers frequently used positive feedback to express approval of the content of learners' messages irrespective of well-formedness. These findings suggest that the corrective reformulations entailed in recasts may not always be perceived by learners because they can be easily overridden by their functional properties in meaning-based classrooms.

Although Lyster's main objective was to investigate "aspects of L2 classroom discourse that may minimize the perceptual salience of recasts and thereby limit their propensity to be noticed as negative evidence" (p. 56), noticing was not directly operationalized in the study. This task was carried over by Egi (2007b), who explored how linguistic targets, length, and number of changes might affect learners' noticing and interpretation of recasts. Recasts of morphosyntactic and lexical errors were provided to learners during task-based activities. Information of their noticing and interpretation of recasts was

gathered with immediate recall and stimulated recall. Results indicated that recasts were occasionally interpreted as responses to content when they were long and substantially different from learners' problematic utterances. In contrast, when recasts were short and closely resembled the original utterances, learners were significantly more likely to attend to the linguistic evidence. The patterns were observed in both morphosyntactic and lexical recasts. These findings suggest that length and number of changes might partially determine the explicitness of recasts and thus affect learners' abilities to interpret them.

A study about learners' interpretation of feedback that has been frequently cited is Mackey et al. (2000) as described in Section 2.4.2.2. The characteristic the researchers focused on was the linguistic target of interactive feedback. Drawing on stimulated recall comments, they found that learners were more accurate in their perception of lexical, semantic, and phonological feedback than morphosyntactic feedback. In a more recent study carried out in Arabic foreign language classrooms, Mackey and other colleagues (2007) again examined learners' interpretation of feedback in relation to its linguistic target. Learners engaged in lessons on a range of linguistic targets (e.g. phonology, morphology/lexis, and syntax) in a number of different ways. After the lessons, both teachers and students participated in a stimulated recall session. Their comments were then analyzed to provide information about whether the learners understood the intentions of the teachers who provided the corrective feedback. The researchers found that learners' perception and teachers' intention about the linguistic target of corrective feedback overlapped the most when the feedback concerned lexis and was provided explicitly.

To sum up, learners' noticing and perception of feedback can be influenced by its characteristics such as length, number of changes, and linguistic targets. Many studies which examined learners' noticing and perception of feedback concentrated on the characteristics of one particular type of feedback or one feature of different types of feedback. It is interesting to explore learners' noticing of feedback in relation to different characteristics across different types of feedback.

2.5.2 Nonlinguistic cues in feedback

A second factor that may affect the effectiveness of feedback is nonlinguistic cues, including paralinguistic cues and extralinguistic cues. In his book which examines literacy from multidimensional and interdisciplinary perspectives, Kucer (2005) gives a clear definition of the two terms in oral language. Paralinguistic cues are features that are part of language but are not linguistic in nature, for example, pitch, intonation, stress, and rhythm. Extralinguistic cues are those that often accompany the use of language but are not part of language, for examples, facial expressions, gestures, physical movements, settings, and objects in the environment. While Kucer classifies nonverbal behavior as extralinguistic cues, not all researchers do so. Davies (2006), for example, categorizes focus-on-form episodes with body language as paralinguistic focus on form. Gullberg and McCafferty (2008) argue that to study gesture is to study the paralinguistic modes of interaction. The word "paralinguistic" as used by these researchers would be the word "extralinguistic" as used by Kucer. The present study adopted Kucer's definition of paralinguistic and extralinguistic cues.

Talking about learners' noticing of recasts, some researchers have suggested that paralinguistic and extralinguistic cues might provide additional communicative clues to learners. Such observations have come from both L1 and L2 acquisition literature. In the L1 acquisition literature, Farrar (1992) suggests that paralinguistic cues might contribute to children's elimination of certain incorrect rules. In the SLA literature, Long (2007) observes that there is some suggestive evidence that subtle prosodic and extralinguistic cues can help with the disambiguation of recasts. Such suggestive evidence can be found in recast studies. Sheen (2004), for example, in a study which examined the similarities and differences among teachers' use of corrective feedback and learners' uptake and repair across four instructional settings found that some teachers in New Zealand ESL and Korean EFL classrooms recast with rising intonation and emphasis, which might have partially increased the salience of recasts and hence learners' noticing of recasts. Loewen and Philp's (2006) study confirmed the effect of intonation and stress. In examining the effect of recasts, they found that declarative intonation and stress, among other characteristics of recasts, predicted successful uptake whereas interrogative intonation was predictive of test scores. As for extralinguistic cues, Sime (2006) explored the meanings that EFL learners attributed to teachers' gestures. Using a stimulated recall protocol, she demonstrated that learners generally attended to teachers' gestures, and the gestures generally helped to clarify verbally expressed meanings. Davies (2006) examined the effect of body language (defined by David as paralinguistic cues) on learners' production of uptake by comparing implicit episodes with or without extralinguistic cues. It was found that episodes with extralinguistic cues tended to result

in more learner uptake than topic continuation while those without extralinguistic cues tended to lead to more topic continuation than learner uptake.

The role of nonlinguistic cues in learning and whether learners attended to them at all, however, are contentious. There is counterevidence too. Carpenter et al. (2006) compared the role of linguistic contexts and nonverbal cues in advanced ESL learners' interpretation of recasts. Two types of videos clips with recasts and repetitions were created: one with both learner utterances and the responses and one with the responses only. Two groups of learners were asked to view the videos and indicate what responses they were hearing. It was found that the utterance-response group was significantly more successful than the response-only group in recognizing recasts, indicating that the immediate discourse context played an important role in learners' perceptions of recasts. The think-aloud protocol from a subset of 14 participants, on the other hand, suggested that among a total of 252 comments there was only one clear reference to the use of a facial expression as a clue to the nature of a recast. The researchers hence concluded that the learners, even those who could not draw on linguistic context, did not use nonverbal cues to interpret recasts. There have also been studies which suggest that even when learners attend to nonlinguistic cues, the effect is not always positive. In a descriptive study, Faraco and Kida (2008) found that while teachers' gaze worked at the dynamic level of interaction and their gestures functioned like a metalinguistic gloss, visual information could also lead to misunderstandings. For example, when teachers softened their corrective moves by breaking their gaze toward a learner or when they produced the

same gesture upon both the repetition and correction of a learner mistake, learners might not understand the corrective nature of the teacher response.

Altogether, there is no agreement on the role of paralinguistic and extralinguistic cues in learners' noticing of feedback and hence the effectiveness of feedback. Although there is suggestive evidence that it might affect noticing and learning, more research is needed to explore it further.

2.5.3 Metalanguage in feedback

There are both proponents and opponents of metalanguage use in the language classroom. Many researchers are strongly against terminology use in language teaching. Corder (1973), for example, believes that the level of abstraction created by terminology adds to learners' burdens. Halliwell (1993) speculates that terminology can only enable learners to talk about language rather than to use it for communication. Unlike Coder and Halliwell, other researchers have assigned terminology to a more important role. Berman (1979) argues that grammatical labels provide a shortcut to all kinds of devious circumlocutions. Faech (1985) observes that metalanguage is a monitor for language production and serves as an important heuristic tool for learners to elicit information about a language. In the call for studies of explicit language talk, Borg (1998, 1999) conducted two empirical studies which examined teachers' use of metalanguage and factors that impinged on their use of metalanguage. From his findings, Borg (1998) concluded that metalanguage can make students aware of the language system they are

learning and how it compares with their first language. Moreover, it also helps students become aware of their own linguistic needs.

In recent years, many researchers have referred to metalinguistic feedback in their studies about the effectiveness of feedback (e.g., Lyster & Ranta, 1997; Panova & Lyster, 2002; Sheen, 2004), which often involves the use of metalanguage. Metalinguistic feedback has been found to be more successful than recasts in eliciting learner uptake and repair in some studies. Nevertheless, metalanguage is not the (only) focus of investigation in these studies. One study which did focus on this aspect of interactive feedback is Basturkmen et al.'s (2002) examination of teachers' and students' use of metalanguage and its relationship with learner uptake.

The study was carried out in two communicative classrooms, one of the intermediate level and one of the pre-intermediate level. The researchers found that the use of metalanguage and the occurrence of uptake varied in different types of focus-on-form episodes. The two occurred together in 50.3% of student-initiated episodes but only over 5% of teacher-initiated and reactive episodes. Further analysis revealed that the relationship between the use of metalanguage and successful uptake also differed according to different types of focus-on-form episodes. While metalanguage and successful uptake occurred together in over 44% of student-initiated episodes, the percentages of such occurrence in teacher-initiated and reactive episodes were as low as 27.3% and 9.5% respectively. Compared with discussions about the use of grammatical terminology, Basturkmen et al's study looked into a broader scope of metalanguage by

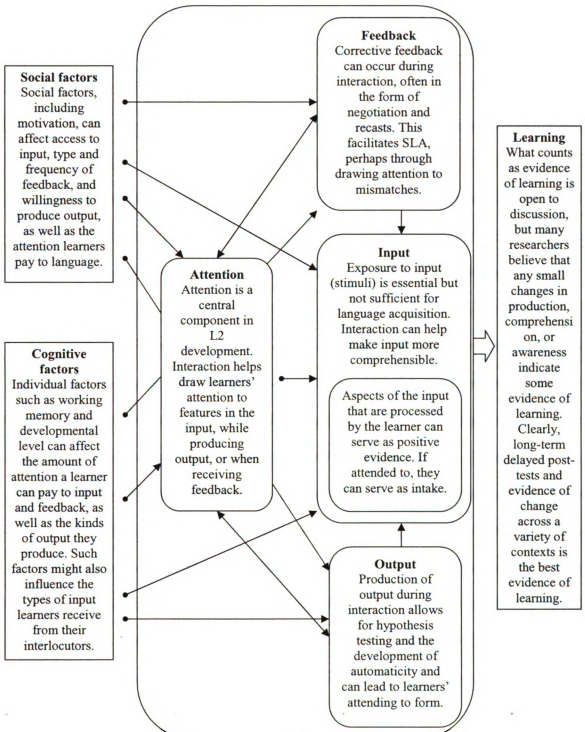
situating its use in a more inclusive environment, focus-on-form episodes, which might contain not only grammar but also other language units such as vocabulary and pronunciation.

The uptake studies mentioned above provide indicative evidence that metalanguage use may affect learners' noticing of feedback and the effectiveness of feedback. However, there is no general conclusion due to the scarcity of research in this area. More studies are needed to provide empirical evidence. Metalanguage use is therefore another aspect of feedback the current study aimed to explore.

2.6 Summary

This literature review covers some of the most important components of SLA: input, output, interaction, feedback, and attention and noticing. In the teaching and learning processes, these components are intertwined. Through the negotiation of meaning in interaction, learners produce the target language and receive feedback from the interlocutor. When the input in the feedback is noticed and attended to, learners may improve their language production, which in turn may affect interaction and the ultimate learning. Recently, Mackey and Polio (2009) nicely present these intertwining tenets of the interaction approach in a diagram (Figure 1). In addition to cognitive factors, they include social factors such as motivation. Limited by time and resources, the current study only focused on learners' noticing of feedback and its effect on learning.

Figure 1 Main tenets of the interaction approach



(Mackey & Polio, 2009, p. 5)

2.7 Research questions

The current study explores learners' noticing of teacher feedback and the effectiveness of teacher feedback in relation to the characteristics of feedback episodes, nonlinguistic cues, and metalanguage. The research questions are:

1. Do the characteristics of teacher feedback episodes affect learners' noticing and learning?
2. Do the nonlinguistic cues in teacher feedback affect learners' noticing and learning?
3. Does the metalanguage in teacher feedback affect learners' noticing and learning?

In many of the studies reviewed above, only corrective feedback was examined. In this study, however, teachers' feedback about questions initiated by learners or by teachers themselves was also investigated. A feedback episode is therefore defined as a sequence beginning with an erroneous utterance, *a query by the learner, or a question by the teacher*, followed by feedback provided by the teacher, and ending with the learner's reaction to the teacher's feedback where applicable (adapted from Lyster and Ranta, 1997). It is important to point out that the term "feedback episode" as defined here is not exactly the same as the term "focus-on-form episode" as used in the studies reviewed above. Focus on form in those studies occurred in a discourse that was primarily meaning centered; the attention to language form was incidental and transitory; and a variety of language forms could be attended to in a single lesson (Ellis et al., 2001a, 2001b). In the current study, apart from such focus on form activities, there were also lessons where pre-selected forms were taught in communicative activities, for example, talking in pairs

about one's summer plan with the "be going to" structure; describing a timeline in groups to practice past tense; and playing a game as a whole class using adverbial clauses. On such occasions, the discourse is not strictly meaning centered but more form oriented; the attention to language form is preplanned and proactive rather than incidental and transitory; and only a limited rather than a large number of language forms might be attended to in a single lesson. In the effort to disentangle the myth of focus on form, Ellis et al. name such attention to language form as the stretched version of Long's (1991) focus on form. The current study does not concern itself so much with the nature of the discourse. Any language episode involving teacher feedback, as long as there was an observable learner problem, was considered as a feedback episode. To put it simply, the term "feedback episode" in this study embraces both focus on form in its original definition and focus on form in its stretched definition. Revolving around this notion of feedback episode, the next five chapters detail the design and findings of the current study.

CHAPTER 3

METHODOLOGY

3.1 Introduction

This chapter outlines the methodological issues of the study. These include the settings of the study, the general information of the participants, the instruments and procedures of data collection, and the schemes of data analysis. In addition, the results of reliability check are presented and the statistical analysis of the data is introduced.

3.2 Teaching context

The study took place in an intensive English program at a large university in the United States. The program is designed to help students learn communication and academic skills that are necessary in daily life and academic work. Students are put into different course levels according to the results of placement tests. The lessons at each level cover grammar, reading, writing, speaking, and/or listening skills. The courses offered are based on a communicative curriculum. In this study, four integrated listening and speaking classes, two integrated reading and writing classes, two academic reading classes, and one grammar class were observed. In each class, learners were involved in all kind of communicative activities, discussing and comparing answers in groups, summarizing texts and stories in pairs, doing information gap activities, to just name a few. Even the grammar class, as the teacher put it: "is surprisingly communicative even though the students need a lot of help" (Observation note, April 14, 2009).

3.3 Participants

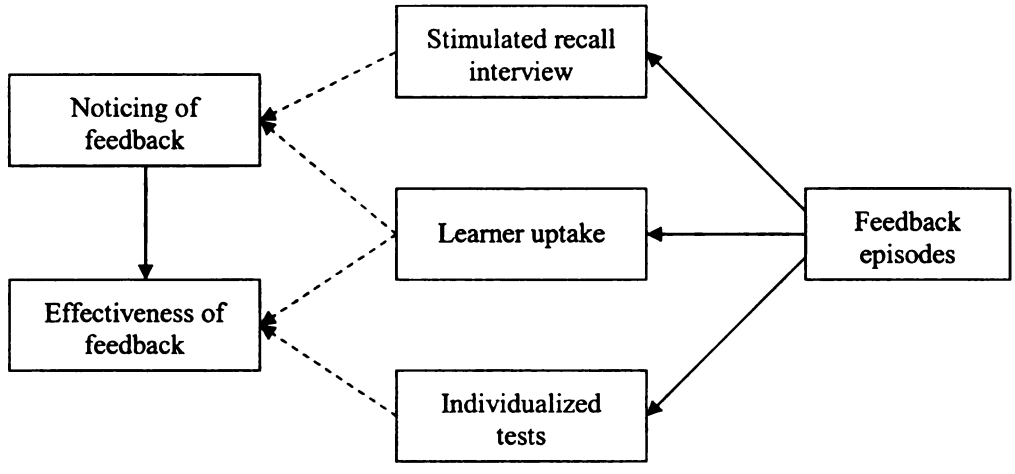
A total of eight teachers and nine classes participated in the study, with one teacher teaching two parallel classes. All the teachers except one have English as their native language. The one exception is a teacher whose first language is Spanish. However, this teacher grew up and went to school in the US. Her English is therefore like that of a native speaker. Among the eight instructors, six are females and two are males. Their teaching experience varies from 1.5 years to 37 years, with an average 10.3 years. At the time of the research, four of the teachers had a MA degree in TESOL; two held an elementary and middle school teaching certificate; one had a teaching English as a foreign language certificate; and still another one was working on her MA degree.

A total of 117 students signed the consent form. Fifteen students who chose not to do so were excluded from the study. The 117 students, with an average age of 20.4, are from a variety of L1 backgrounds, including Chinese, Arabic, Korean, and Japanese. All of them had studied English before they came to the US, with an average 7.3 years of previous English study. The majority of students had arrived in the US not too long, with an average 5.7 months of stay. Only a few had been in the country for 12 months or more. Most of the students enrolled into the intensive English program to improve their English either to qualify for admission to regular American academic institutions or to return to their home country for further career development.

3.4 Instruments

According to the research questions, two variables were measured in the present study: learners' noticing of teacher feedback and the effectiveness of the feedback. In response to the call for triangulating data collection methods, each variable was operationalized with two measures. As shown in Figure 2, the noticing of feedback was examined with learners' stimulated recall comments and the effectiveness of feedback was assessed with individualized tests. Learner uptake was used as an additional measurement for both noticing and learning. All three measures were developed from teacher feedback episodes.

Figure 2 Measurement of the noticing and effect of teacher feedback



Accordingly, three instruments were used in the study: classroom observations, stimulated recall interviews, and language tests. Classroom observations provided data for learner uptake as well as the characteristics of feedback episodes, the occurrence of nonlinguistic cues, and teachers' metalanguage use. The stimulated recall interviews and

language tests were also based on the observations. Below is a detailed illustration of the three instruments.

3.4.1 Observations

Classroom observations, as shown in the studies reviewed in Chapter 2 (e.g., Basturkmen et al., 2002; Loewen & Philp, 2006; Lyster & Ranta, 1997; Oliver & Mackey, 2003), are a data collection technique widely adopted in SLA. As Gass and Mackey (2007) put it, observations "allow researchers to gather detailed data on the events, interactions, and patterns of language use within particular foreign and second language classroom settings" (p. 165). In the present study, four to six observations, depending on teachers and students' consent, were made in each classroom. During the observations, I served as a non-participant observer and did not participate in any of the classroom activities. My job was to monitor the recording instruments and to take notes. A Sony digital voice recorder with a clip-on microphone attached to the teacher was used to record the oral exchanges in the classrooms. As Loewen (2004) notes, this arrangement can record all teacher-student interactions, whether the students interact with the teacher as a whole class, in small groups, or one on one. However, it does not capture nonverbal interactions, which is an important component I intended to examine. To tackle this problem, in addition to the voice recorder, a Sony high definition video camera with a built-in speaker was set up wherever the teacher suggested was more convenient and less obtrusive.

3.4.2 Stimulated recall interviews

Stimulated recall interviews, along with uptake from the observation data, were major tools to collect noticing data. One to two recall interviews were conducted in each class depending on the number of observations that were allowed by the teacher and learners. In each class a minimum of four students were invited to participate in this session, depending on the circumstances and time frame. In order to mitigate the reactive effect of factors such as double-input exposure and verbalization, data from the interview observation(s) were not reused for testing. That is, test items were developed from other observations. In line with the recency rule proposed by Gass and Mackey (2000, 2007) and to reduce the non-veridicality caused by memory decay, the interviews were conducted within three days after the observation(s). It has been suggested that recall may be more successful when the stimuli are presented in the same way they were presented in the previous task so that the sequence of the cues is most likely to be consistent with memory indexes in long-term memory (Egi, 2004). For this reason, it would be best to have learners view the recordings of whole or big chunks of class periods. However, due to time constraints, and also in the effort not to burden learners with too much extra work, excerpts of the video recordings of classroom interactions (i.e. feedback episodes) were used to serve as recall cues. To examine learners' noticing of teacher feedback in different circumstances, clips from a variety of settings were selected, some from interactions between teachers and individual students, some from group work, and some from whole class discussions. Only excerpts of both good audio and visual quality were selected. It was hoped that high quality stimuli would be able to help learners retrieve their memory more easily and more thoroughly, thus revealing more about learners'

cognitive processes and enhancing the veridicality of the noticing data. Because one of the objectives of the study is to examine the effect of nonlinguistic cues on learners' noticing of teacher feedback and yet nonlinguistic cues did not occur frequently in every observation, episodes with nonlinguistic cues received special attention in the excerpt selection process and were selected for use as long as they are of good quality. However, this is not an issue for teachers' metalanguage use now that metalinguistic terms occurred much more frequently than nonlinguistic cues.

After the excerpts were selected, learners who were engaged in these episodes were invited to view the video clips and report what they were thinking. On some occasions observers of an episode[1] were also asked to participate in the interview session. This generally happened with episodes from group work or whole class discussion. It was hoped that observers' reports could help reveal learners' noticing of teacher feedback when they were not directly engaged in teacher-student interactions. To reduce interviewer effect, I did a pre-viewing debriefing before each interview, explaining to learners, with specific examples, that they should report their then-thinking, not their now-thinking, and that they did not need to worry about what I thought of their performance or reports. They were also informed that they could stop the tape whenever they had a comment to make. During the interview, at points where there were communication breakdowns, for example, when the learners looked confused, when there

[1] The observer of a feedback episode refers to a learner who is not directly involved in a teacher-student interaction but still experiences it. For example, in a whole class discussion, a learner might initiate a question which other learners turn out not to understand either. As the teacher gives feedback to the question-initiator, other learners hear it too and hence are observers of the episode.

was a long pause in the conversation, and when the learners stuttered, I also stopped the tape to ask them what they were thinking at the moment.

Ericsson and Simon (1993) suggest that participants should only be asked to verbalize information that has been attended to. Only such information can be encoded in memory and available for verbal reports. They also recommend that the instructions for recall should avoid a request for reasoning and elaboration from respondents. Such requests may result in participants altering their thoughts in order to generate the information requested. For these reasons, I only asked questions such as "What were you thinking then?" or "Any thoughts there?" not questions such as "Why were you thinking so?" or "Could you tell me more about that?" When learners said they were not thinking of anything, I did not push them to think further.

One thing I noticed in the recall session is that with English as a second language, some learners, especially those at a lower proficiency level, had difficulty expressing themselves on some occasions. For example, a few learners were not able to remember a particular word and so asked "How do you say...?" or "What's the opposite word of ...?" In this case, I gave them some language help as I saw necessary. Meanwhile, I took caution not to give them extra information. I hoped this approach would help prevent the unnecessary loss of valuable data due to learners' limited proficiency level. Another thing that stood out in the recall session is that about 60% of the learners who did the interview reported that their native language is Chinese, which happens to be my native language too. To reduce non-veridicality to the largest extent, and also given that the purpose of

the recall interview is to obtain noticing data, not to test learners' language skills, these learners were allowed to speak Chinese if they wanted to. It was hoped that this would give them a chance to express their thoughts more thoroughly and more accurately.

3.4.3 Language testing

Loewen (2002) developed a battery of individualized tests in his study on the effectiveness of incidental focus on form. Given the careful design of his study, his testing method was adopted in the current study. Altogether, there were six types of test items: 1) supply meaning, 2) supply word/phrase, 3) correction, 4) pronunciation, 5) spelling, and 6) supply information.

In Supply Meaning, learners were asked to provide the meaning or definition for a word or phrase. For example:

(6) Please tell me the meaning of the word: Lower.

"Lower" had been discussed in a feedback episode where the learner asked the teacher what the word meant and the teacher explained it with synonyms and gestures.

In Supply Word/Phrase, I provided the meaning of a word/phrase and learners were asked to give the word/phrase. For example:

(7) When the quality of the soil gets worse and worse, or there is less and less nutrition in the soil, we say the soil is losing F_____.

Here the expected answer is "fertility", which had been discussed in a feedback episode. In this type of tests, the first letter of the word was given in order to ensure that learners would not give another equally correct word.

In Correction, I read incorrect sentences to learners and explicitly asked learners to correct them. For example:

(8) The following sentence is incorrect or inappropriate. Please listen carefully and tell me how you could make it better.
During the spring break, I will go to many place in California.

The inaccurate form "many place" had been corrected in a teacher-student interaction. The learner's task was to listen to the sentence and then correct it.

In pronunciation tests, learners were asked to read a sentence that contained a targeted word and then read the word in isolation. For example:

(9) Please read aloud the following sentence and word.
The word passionate means very eager or exited.
Eager

Here the sentence is very similar to the one the learner had produced in a feedback episode, where the word "eager" was pronounced incorrectly and was recast by the teacher.

In spelling tests, learners were asked to orally spell a word that had received attention in a feedback episode. For example:

(10) Please spell the following word: sibling

In a previous teacher-student interaction, the learner had received feedback from the teacher about the spelling of the word "sibling".

Items which required learners to supply information were often associated with feedback episodes where students asked metalinguistic questions. The prompts for such items might be questions such as "Is the word 'lower' positive or negative?" This question was developed from an episode where the learner consulted the teacher whether lower is positive or negative.

Similar to Loewen's study, there was no pretest but an immediate test and a delayed test based on different observations. However, three changes were made as I saw necessary in the context of the current study. In Loewen's study, only students who initiated a question were tested on a particular linguistic feature. In the current study, I noticed that in many cases when learners were discussing questions, they decided together that they needed the teacher's help. Even though only one student initiated a question, other group members would listen and respond to the teacher. For example:

(11) Maj: T, we don't understand about me too.
T: Oh me too is useful. So what are you going to do this summer? Are you going to do anything the same?
Maj: (Shakes head)
T: Nothing? What about this one? (Points at Muh's paper) Are you going to study English?
Maji: Yeah, but I didn't write it.
T: Oh ok, but still in a conversation, if he said I'm going to study English, you can say [me too].
Muh: [Me too]. Yeah. (Nods)
T: Yeah. It's true you didn't write it, but that's ok, so me too. Understand?
Maj: Yeah I understand. (Nods)

In this example, the question initiator uses the word "we" when asking the question, suggesting that both he and his partner need help with the phrase "me too." Although Muh is not the one who asks the question, he listens to the teacher's explanation and responds to her by saying "me too" and nodding. This learner therefore also experiences

the feedback episode. This is also the case when the teacher initiated a question and more than one student showed or said they did not know the answer or when a group made the same mistake and the teacher corrected them all. On such occasions, the student(s) who was (were) not the initiator of a question was (were) also tested on the same item.

In the case of supplying meaning and spelling, Loewen did not provide a full sentence containing the targeted word or phrase to test takers. This is probably a good way to prevent learners from doing guess work according to the context and more precisely tap into their implicit knowledge about the targeted linguistic features. As I attempted to ask learners such questions, however, it was found that learners sometimes talked about another word or phrase as a result of homophones or similar words. For example:

(12) Please spell the following word: phase

Among seven learners who were asked this question, four gave the word "face" without hesitation. In order to make the prompt clearer, it was decided that a sentence containing the targeted word or phrase could be provided when necessary. The prompt of the above example thus would become:

They're in the second phase of love. Tell me the meaning of the word: phase.

By providing the sentence, learners would hopefully be able to see the "phase" I was asking about is not the body part "face", which would not make sense in this context. In order to minimize additional comprehension problems, the sentence was generally very similar to, sometimes the same as, the one that occurred in the original feedback episode. On the other hand, it was realized that the addition of such sentences, especially in the case of supplying meaning, could easily increase the chance of wild guesses. To address

this problem, the sentence was made as short as possible and contained as little additional information as possible.

In order to ensure the reliability and validity of his tests, Loewen made a great effort to design test items that reflect the modality of focus-on-form episodes in real-time classroom interactions (e.g., all test types were oral rather than written). To further address the problems incurred by the mismatch between test and non-test situations, an additional change was made in line with one of Ellis' (2005) most important criteria for the assessment of implicit and explicit knowledge, timing. As Ellis observes (see 2.3.2.1), a test task that taps into a learner's implicit knowledge does not allow the learner to spend an extended time planning a response. This proved true in Ellis' (2005) psychometric study. Following this rule, learners were encouraged to answer the questions as quickly as they could. In addition, the time learners spent answering each question was recorded. With that said, however, there is no guarantee that the tests indeed or merely measured learners' implicit knowledge. As Ellis observes, "even if task conditions that inclined learners to use one type of knowledge in preference to the other could be identified, it would be impossible to construct tasks that would provide pure measures of the two types of knowledge" (p. 153).

3.5 Procedures

Before starting formal observations in a class, I distributed and collected the consent form as well as a small background survey in the class. With the teacher's consent, I then stayed in the classroom for an hour or two to get familiar with the teacher's teaching style

and to select possible positions to put the video camera. This also gave me the opportunity to talk to learners and to try to establish a rapport with them, which could hopefully help increase the ecological validity of the study.

Data collection in each class took three to four weeks, depending on how many times teachers and students agreed to have me with them. There was no fixed schedule due to the complexity of particular contexts. Sometimes learners spent all their class time taking a unit test. Sometimes the majority of class time was devoted to silent writing. Sometimes teachers were out of town for a conference. On the other hand, both the tests and stimulated recall interviews, especially the latter, were highly time sensitive. Consequently, when anything unexpected happened, the observation and the test or interview based on it would be cancelled and new arrangements would be made accordingly. Because the tests and interviews were based on different observations, such adjustments should not severely affect the results.

Generally, when there were six observations, Observations 1 and 2 would serve as the basis of the delayed test, Observations 3 and 4 would serve as the basis of the stimulated recall interviews, and Observations 5 and 6 would serve as the basis of the immediate test. When there were four to five observations, the delayed test would be developed from Observation 1, the recall cues for the interviews would be selected from Observation 2, and the immediate test would be developed from Observation 3. Observations 4 and 5 would supply additional material for the tests and interviews and I decided whether more

test items were needed and whether more recall interviews should be conducted according to what I had already.

The immediate test was administered one to three days after the observation from which it was developed while the delayed test was administered 14 to 16 days after the observation from which it was developed. As Loewen (2002) notes, giving a range of days for the immediate test and delayed test is to allow some leeway for administering tests. For example, if students cannot take the immediate test on one day, they can take it in the other two days. Ideally, the stimulated recall interviews should be conducted right after an observation. However, this turned out to be very difficult in the current study. For one thing, it was not quite possible for me to identify feedback episodes and select recall cues during the observation: it took quite a while even just to import the video to a laptop and to convert it to an easily viewable and editable format. For another thing, learners often had other classes to attend right after an observation or they were often too tired after a long day and were reluctant to do the interview. More often than not, it was even not easy to arrange a meeting time in the following day(s). For such reasons, it was decided that a three-day range was also necessary for the stimulated recall interviews. Meanwhile, every effort was made to shorten the time between the observation and the interview. The average time span is about 48 hours. If a learner for some reason was not able to take the tests or to do the interviews within the three-day deadline, the tests and interviews would be cancelled.

3.6 Data analysis

Three kinds of data were analyzed: classroom observations, stimulated recall comments, and test results. The observation data were coded in terms of the characteristics of feedback episodes, the nonlinguistic cues in teacher feedback, the metalanguage in teacher feedback, and learner uptake. The stimulated recall comments were coded according to their relevancy to language in general, the target features in the stimuli, and teacher feedback in the stimuli. Finally, learner responses in the immediate test and delayed test were coded as correct, incorrect, or partially correct. After all data were coded, a random sample of about 10% of each data set was coded by a second rater and chi-square tests were performed where necessary and possible.

3.6.1 The coding of observation data

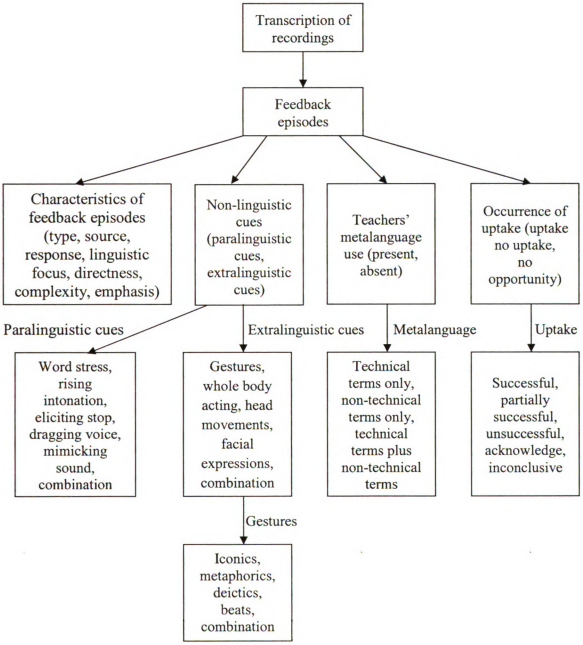
The first step of coding the observation data is to identify all episodes that contain teacher feedback. These feedback episodes were then coded according to their characteristics, the nonlinguistic cues teachers used, the metalanguage teachers used, and learner responses to the feedback. Some of these variables were then further analyzed. The coding process is shown in Figure 3.

3.6.1.1 The coding of the characteristics of feedback episodes

As shown in the literature review, R. Ellis and colleagues (e.g., Ellis et al., 2001a; 2001b; Loewen, 2003, 2004, 2005) have developed a coding scheme to examine the characteristics of focus-on-form episodes. A careful look at these characteristics (see Table 1) reveals that all of them revolve around feedback. It is no exaggeration to say that

feedback is at the core of focus on form. The characteristics of focus-on-form episodes therefore can also apply to feedback episodes. In the current study, the coding scheme in Loewen (2004) was adopted to examine the general characteristics of feedback episodes.

Figure 3 Coding process of the observation data



As illustrated in Table 1, seven characteristics were examined: type, source, response, linguistic focus, directness, complexity, and emphasis. In his study, Loewen (2004) included uptake and successful uptake as two characteristics of focus on form. In the present study, these two variables were coded separately because they were important tools to examine both learners' noticing of teacher feedback and their language improvement.

Table 1 Coding scheme for the characteristics of feedback episodes

Characteristics	Definition	Categories
Type	How the feedback episode is initiated	*Teacher-initiated*: teacher starts a question *Student-initiated*: student starts a query *Reactive*: teacher responds to erroneous utterance
Source	What causes the feedback episode	*Code*: inaccurate use of language forms without communication problems *Message*: problems understanding meaning *Semantic:* negotiation involves meaning of a word/phrase but there's no communication breakdown
Response	How the feedback is given	*Provide*: teacher recasts or provides information about language form *Elicit*: teacher strategically draws out language information from student
Linguistic focus	What language aspect is targeted at	*Vocabulary* *Grammar* *Pronunciation* *Spelling*
Directness	How explicit the feedback is	*Direct*: explicit (e.g., metalingual explanation) *Indirect*: implicit (e.g., recasts)
Complexity	How many feedback moves there are	*Simple*: one feedback move *Complex*: more than one feedback move
Emphasis	Combination of complexity and directness	*Light:* Indirect and simple *Heavy:* Direct and/or complex

Adapted from Loewen (2004, p. 165-166)

Type

Type refers to how a feedback episode is initiated. This characteristic is classified into three categories: teacher-initiated, student-initiated, and reactive. In a teacher-initiated episode, the teacher starts a language question and gives learners feedback if they can not provide an answer to the question or provide an incorrect answer. In a student-initiated episode, the learner starts a query about a language feature and the teacher gives feedback to the learner. In a reactive episode, the learner produces an erroneous utterance and the teacher responds to the error. Below are three examples which show the three types of episodes.

(13) T: About the football game. It's coed. Do you know what coed means?
Class: No.
T: It means boys and girls together.
Class: Ah.

(14) Lofa: What to say when there are three semesters?
T: So if there were three, they would be called trimesters.
Lofa: Yeah. (Smiles)

(15) T: Anybody have something different? Le?
Le: She don't want children.
T: Ok, the first thing is she doesn't want children. (Writes on board)
Le: (No more response)

Here the first episode is initiated by the teacher asking the class about the meaning of the word "coed" while the second episode is initiated by a learner asking the teacher about the word to describe three semesters. These two episodes are teacher-initiated and student-initiated respectively. In the third example, the learner uses the wrong form of the helping verb "do" and the teacher recasts him. The teacher's feedback is a reaction to the learner's response. It is therefore a reactive episode.

Source

This characteristic describes what causes a feedback episode. It is put into three categories: code, message, and semantic. A code-related episode results from the inaccurate use of language forms without communication problems. The purpose of teacher feedback is to increase learners' language accuracy. A message-related episode results from the inaccurate use of language forms with communication problems. The purpose of the episode is to solve comprehension problems. In a semantic episode, there may not be a communication breakdown; the learners might have understood the gist of the discourse but wants to know the meaning or usage of a specific word or phrase. Below are three examples that show the three types of sources.

(16) Vah: And the two couple, one couple, they're enjoying, they're enjoying their love.
T: Ok.
Vah: Yeah. They close their eyes. So they're blind.
T: Ok, they're blind-, they're blindly in love, ok.
Vah: They're blindly in love.

(17) Happ: Fashion trend. ("Fashion" sounds like "fishing")
T: Huh?
Happ: Fashion. ("Fashion" still sounds like "fishing")
T: Fishing? (Hand gestures fishing)
Happ: Fashion. (A little bit clearer)
T: Fashion, fashion trend, yeah. Eh ladies you know about fashion trends?

(18) T: Cute.
Tai: It's me.
Vah: You!
Class: (Laugh)
T: Usually they say girls are cute, boys are, babies are cute, but boys are usually not cute.
Vah: Pet, pet.
T: Pets are cute. So the way they look. All right?
Ara: Yeah.
Tai: (No response)

The three episodes are code-related, message-related, and semantic respectively. In the first episode, the learner is describing a picture to the teacher. There does not seem to be a communication problem. The teacher recasts the learner to help him express his ideas in a better way. Different from the teacher in the first episode, the teacher in the second episode does seem to have difficulty understanding what the learner is saying because of his problematic pronunciation of the word "fashion". The teacher attempts to solve the comprehension problem with clarification requests. In the third episode, the learner uses the word "cute" to describe himself, which the teacher explains is inappropriate. There is no communication breakdown. The teacher's feedback is a response to the fact that the learner does not have a full understanding of how to use the word "cute".

Response

Response refers to how the teacher's feedback is given. It is classified into two categories: provide and elicit. When the response type if provide, the teacher recasts the learner's erroneous utterance or provides information about a language feature. When the response type is elicit, the teacher strategically draws out language information from the learner. Below are two examples which show the two types of response.

(19) Yiw: It's almost a pa-, pa- (Not being able to pronounce the word "passionate")
T: Passionate.
Yiw: Passionate about his girlfriend.

(20) Jacq: He thinks studying English is not good enough, so he went to here to study English in the ELC.
T: He thinkS (Stresses "s") studying English is not good enough back home? (Rising intonation) That's why he came here?
Jacq: He thought.

In the first episode, the learner can not pronounce the word "passionate". The teacher simply models it for her. In the second episode, the learner uses the wrong tense of the word "think". Instead of providing the correct form, the teacher repeats the learner's erroneous utterance in a rising intonation, stresses the third person "s" in "thinks", and asks a question in the past tense. With such strategies, he successfully draws out the correct form of "think" from the learner.

Linguistic focus

Linguistic focus refers to what language aspect is targeted at in a feedback episode. It can be vocabulary, grammar, pronunciation, or spelling. In vocabulary episodes, the linguistic focus can be the meaning of words, phrases, idioms, and sentences, word choice, non-target derivations of words, pragmatic aspects of words or phrases such as their appropriateness in specific social contexts, whether a word or phrase is positive or negative, etc. In grammar episodes, the linguistic focus can be determiners, prepositions, pronouns, word order, tense, verb morphology, auxiliaries, subject-verb agreement, plurals, negation, question formation, plural –s, sentence construction, comparative and superlative, part of speech, etc. In pronunciation episodes, the linguistic focus may involve segmental and superasegmental aspects of the phonological system such as the pronunciation of words, stress, and intonation. In spelling episodes, the linguistic focus is the orthographic form of words. Below are four examples which show the four types of linguistic focus.

(21) Lyb: Like loosen the soil. Remove the soil. (Hand gestures loosening)
T: Move the soil?
Lyb: Loosen like eh exchange. (Hand gestures exchanging)
T: Not...Mm, yeah they need to tear the soil. (Hand gestures tearing)

Lyb: Yes.

(22) Fa: They came back in the evening and noticed that the milk was disappeared.
T: The milk HAD disappeared. (Stresses "had")
Fa: Had disappeared.

(23) Gra: But they stopped dating because there was no chemistry- /'tsemistri/
T: Chemistry.
Gra: Chemistry between them. (/'kemistri/)

(24) T: Jealous, to be envious.
Ara: How to spell it?
T: E-N-V-I-O-U-S.
Ara: N-V. (Writes on paper)
T: V-I-O-U-S.
Ara: (Finishes writing on paper without saying anything more)

In the first episode, the teacher and the learner are seeking the right word to describe an action. In the second episode, the teacher corrects the learner's inappropriate use of a helping verb. In the third episode, the teacher is helping the learner with the pronunciation of a word. In the fourth episode, the teacher and the learner are talking about the spelling of the word "envious". Accordingly, the linguistic focus of the four episodes is vocabulary, grammar, pronunciation, and spelling respectively.

Directness

This characteristic describes how explicit teacher feedback is. There are two categories in it: direct and indirect. Direct feedback is explicit, for example, explicit correction and metalinguial explanation. Indirect feedback is implicit, for example, recasts, clarification requests, and repetitions. Below are two examples for direct feedback and indirect feedback.

(25) Zhou: What does that mean? The first word.
T: Yes, eh testosterone. That's a male hormone. That's eh, something that's inside your body that the men have, not the women. Just men.

Zhou: Oh.

(26) Mei: (Talking about what is in a picture) And gas masker, masker.
T: Yeah, gas masks.
Mei: Masks.

In the first example, the teacher explicitly explains the meaning of a word the learner asks about. The feedback is therefore direct. In the second example, the teacher corrects the learner's erroneous production by recasting him. The feedback is therefore indirect.

Complexity

This characteristic involves the number of teacher feedback moves in a feedback episode. There are two categories in it: simple and complex. When there is only one feedback move by the teacher, it is simple. When there is more than one feedback move by the teacher, it is complex. Below are two examples showing the complexity of feedback episodes.

(27) Hasa: Eh can you spell pursue please?
T: P-U-R-S-U-E, pursue.
Hasa: P-U-R-S-U-E.

(28) Xin: What is interminal?
T: Interminal?
Xin: Yeah.
Lo: Not a word.
T: No. That's not a word. (Shakes head)
Xin: Means forever.
T: Interminal?
Xin: Eterminal.
T: Eternal?
Xin: Yeah, eternal.

In the first example, there is only one feedback move by the teacher. In the second example, there are four. Accordingly, the first episode was coded as simple while the second one was coded as complex.

Emphasis

The last characteristic is a combination of complexity and directness. There are two categories in it: light and heavy. Light episodes are indirect AND simple while heavy episodes are direct OR complex or both. Below are three examples illustrating whether an episode is light or heavy.

(29) T: What kind of work do they do?
Casu: A lawyer.
T: Yeah, they're lawyers. [And]...
Casu: [And] they took whatever cases.

(30) Lofa: Can I say too many garbages?
T: Too much garbage.
Lofa: Too much garbage.
T: It's not a count noun, so you can't say garbages. Just garbage.
Lofa: Too much garbage.
T: Yeah.

(31) Rouf: Which form of balance should I use?
T: Eh you want a noun. Something that's not balanced. If something is not balanced, it's out of balance, it's imbalanced, it's unbalanced. That's ok. Global words, word forms.
Rouf: (Nods)

In the first example, the teacher corrects the learner by recasting her inappropriate use of the word "lawyer". There is only one feedback move and it is implicit. This episode was therefore coded as light. In the second example, the teacher provides a direct answer to the learner's question and there is more than one feedback move. This episode was therefore coded as heavy. In the third example, the teacher explicitly explains the usage of the word "balance". Although there is only one feedback move, it is direct. This episode was therefore also coded as heavy.

The two examples in Table 2 illustrate how the same feedback episode was coded according to the seven different characteristics. In the first episode, the teacher asks a question about the word "generate" and explains it when learners' responses indicate that they do not understand it. This episode was categorized as message-related because it involves the meaning of a word with an apparent comprehension problem. It is teacher-initiated because it is the teacher who starts the question. The feedback type is provide because the teacher directly tells learners what the word means. The linguistic focus is vocabulary because it involves word meaning. It is direct because the teacher explicitly gives learners the answer. It is complex because there are two feedback moves by the

Table 2 Examples for the coding of the characteristics of feedback episodes

Episodes	Characteristics	Categories
(32) T: And what does generate mean?	Type	Teacher-initiated
Adu: Generate?	Source	Message
Lyb: Oppress, oppression.	Response	Provide
T: It just means make. It can produce.	Linguistic focus	Vocabulary
Adu: Produce.	Directness	Direct
T: That's another way of saying it.	Complexity	Complex
Lyb: Produce.	Emphasis	Heavy
(33) Lai: I was tooking an English class while Daniel was tooking Spanish.	Type	Reactive
	Source	Code
	Response	Elicit
T: Tooking? (Rising intonation)	Linguistic focus	Grammar
Lai: Taking.	Directness	Indirect
	Complexity	Simple
	Emphasis	Light

teacher. It is heavy because the teacher's feedback is both direct and complex. In the second episode, the learner uses the wrong form of the verb "take" but there does not seem to be any communication problem. Instead of explicitly correcting her, the teacher brings out the correct form from the learner by repeating the incorrect form in a rising

intonation with one single move. This episode is therefore reactive, code-related, indirect, simple, and light. The feedback type is elicit, with a focus on grammar.

3.6.1.2 The coding of nonlinguistic cues

As shown in Figure 3, two types of nonlinguistic cues were identified: paralinguistic cues and extralinguistic cues. After all the episodes were classified as being with or without the two types of cues, those with either type were further classified into more specific categories.

The coding of paralinguistic cues

In the fields of speech communication and nonverbal communication, much attention has been given to paralinguistic features. However, as in many other fields, there does not seem to be a unanimous agreement on the definition and boundary of the term. Trager (1958, cited in Jenkins & Parra, 2003), for example, uses the term "paralanguage" to refer to features such as voice set (physiological and physical aspects), voice quality (pitch range, rhythm, articulation, resonance, and speed), and vocalization (nonlexical sounds). Martin (1981), on the other hand, classifies many of these features into the prosodic domain (e.g., stress and intonational pitch direction) and paralinguistic domain (e.g., loudness, pitch, tempo, voice quality, and voice characterizers). Due to the inconsistency of definition, researchers have used the word "paralinguistic" in different ways. Reviewing previous prosody and discourse studies, Shriberg et al. (1998) found that the acoustic-prosodic features covered in these studies include pitch range, intonational contour, declination patterns, utterance duration, preboundary lengthening phenomena,

pause patterns, speaking rate, and energy patterns. Given the specifics of the current study, it was decided that instead of using the coding scheme of a particular researcher, a few features derived from Shriberg et al.'s list would be examined, namely: word stress, rising intonation, eliciting stop, dragging voice, and the combination of the other categories. The combination category was not excluded because the combined effect of various categories might be different from that of a particular category by itself.

Word stress

With this type of paralinguistic cues, the teacher stresses a particular letter, word, or phrase, which is often crucial to learners' understanding of a linguistic feature. For example:

(34) Nass: Could you tell me test score please?
T: MY test score. (Stresses "my") Could you tell me my test score please?
Nass: My test score.

In this example, the learner leaves out the pronoun "my" in his sentence. The teacher recasts him and stresses the word "my" in her feedback to the learner. The learner notices that and adds "my" to his new utterance.

Rising intonation

Rising intonation is mostly used when the teacher is indicating that the learner's utterance is erroneous and wants to elicit the correct form from the learner or to draw the learner's attention to the correct form. For example:

(35) T: What kind of activities are these?
Hasa: Entertainment.
T: They're entertainment. Anything else?
Hasa: To waste their time.

T: To waste their time? (Rising intonation)
Class including Hasa: To spend their time.

In this example, the teacher is leading a discussion after the class watched a sitcom episode where a family is entertaining themselves. One learner chooses the wrong word for spending time. The teacher repeats the inappropriate sentence in a rising intonation. Upon that, the learner realizes his mistake and corrects himself together with other students.

Eliciting stop

In episodes with eliciting stop, the teacher first gives the learner specific hint(s) and then stops on purpose for the learner to provide the complete answer. For example:

(36) T: Now how do you hit somebody with the back of your hands?
Chub and Pyo: (Silence, 3 seconds)
T: S- (Hand gestures slapping someone with back of hand and stops on purpose)
Chub: Whoa, slap?
T: Ok, yeah. (Nods)
Pyo: Ah!

In this example, the learners have difficulty finding the right word to describe the action of hitting somebody with the back of one's hand. The teacher tries to prompt them to give the word "slap" by giving the first letter of the word. With that hint, one learner figures out the right answer and the other acknowledges it.

Dragging voice

Dragging voice is generally used in two cases. On some occasions, the learner produces a problematic utterance and the teacher makes a short utterance such as "well", "um", "eh", and "bababa" with a lengthened ending voice to show his/her disagreement but without

providing any specific hint. On other occasions, the teacher pauses on purpose after saying part of a word or phrase before finishing the whole utterance. The purpose is presumably to draw the learner's attention to what is coming up next. Below are two examples illustrating how the dragging voice technique is used under the two different circumstances.

(37) Lai: I saw Monique at the club last night.
Lin: What did she do there?
T: Mm-- (Drags voice)
Lin: What was she doing there?
T: Yeah, what was she doing there?

(38) T: Climate change. Ok, what kind of climate change?
Ru: Drough. (/drəut/)
T: Dr-- (drags voice) --ought. Say it. Drought.
Class including Ru: Drought.

In the first example, the learner uses the wrong tense for the verb "do". The teacher simply says mm- in a dragging voice without any specific hint. Upon that the learner realizes his error and gives the correct form in no time. In the second example, the learner mispronounces the vowel in the word "drought". When giving her feedback, the teacher drags his voice after pronouncing the consonant cluster /dr/. With the teacher's encouragement, the learner repeats it correctly in the choral response.

Mimicking sound

With this type of paralinguistic cues, the teacher mimicks the sound of a thing or the noise from an action. For example:

(39) T: Siren.
Hasa: No, no. I don't know.
T: Police. (Imitates siren)
Class: Oh.
Chen: Fire alarm.

T: Like the fire alarm, police siren.
Sa: Don't do that.
T: Don't do that?
Sa: Don't say that.
T: Don't say that? Why?
Yiw: We have a fire alarm last night.

(40) Tal: What is splash?
T: Splash? Psh- (Makes splashing noise and hand gestures water splashing) Water splashes.
Tal: Oh. (Writes on paper)

In the first example, the teacher imitates the sound of a siren, a thing. In the second example, the teacher imitates the sound of water splashing, an action. In both episodes, a real sound is mimicked.

Combination

On many occasions, the teacher uses more than one paralinguistic cue in the same episode, especially rising intonation and word stress. For example:

(41) Joy: They're talking about how many camels does...
T: How many camels DOES? (Stresses "does" and rising intonation)
Joy: Eh do eh [does]
T: [Does]?
Joy: Does, does, does he has.
T: He HAS? (Stresses "has" and rising intonation)
Joy: He have. How many does he have.

In this episode, there are two combinations of rising intonation and word stress. The first one occurs when the learner puts a plural noun and a singular helping verb together (which turns out to be correct when he continues the sentence). The second one occurs when the learner uses the helping verb "does" and yet still uses the third person "has" for the same subject. In both cases, the teacher uses two types of paralinguistic cues.

The coding of extralinguistic cues

The coding of extralinguistic cues is a stepwise procedure. After all episodes with any extralinguistic cues were sorted out, they were further categorized into five types: hand gestures, whole body acting, head movements, facial expressions, and combination. Hand gestures were then further coded as iconics, metaphorics, deictics, beats, or combination. Combination again was included because the use of different types of extralinguistic cues or gestures might have a different effect on learners' noticing and learning from a particular type of extralinguistic cues or gestures.

Discussing his concept of implicit communication, Mehrebian (1972) gives a list of kinesic behaviors: facial expressions, gestures, posture, head movements, gaze, and position. This list provided a basis for the coding scheme for the type of extralinguistic cues in the current study although the terminology is slightly different and not every category in the list was examined.

Gestures

Gestures are the movements of the fingers, the arms, and the hands, "the spontaneous, unwitting, and regular accompaniments of speech" (McNeil, 1992, p. 3). This is one of the most frequently used extralinguistic cues in the classrooms observed. For example:

(42) Smal: What's a pad?
T: What's that?
Smal: The yellow pad. What's that? P-A-D.
T: B-A-D?
Class: P-A-D.
T: P-A-D. Oh pad. So it's a small writing book. Small writing book.
Noa: Like notes.
T: Yeah. Notepad and notebook. It's the same thing. A pad just means like a

rectangular card thing. (Hand gestures "rectangular")
Smal: (No response)

In this episode, the learner asks about the meaning of the word "pad". In addition to verbal explanation, the teacher also uses both hands to form a rectangular, which is the normal shape of a pad.

Whole body acting

With this type of extralinguistic cues, the teacher acts out what is being discussed. The body parts that are used are not just the hand or arm. It often involves the whole body, with the term in its broad sense (i.e., not exactly every part of the body). For example:

(43) T: Perk up.
Class: (Silence, 7 seconds)
T: Perk up. (Pauses, 2 seconds) Ok, students, if you, if you're like this...(Acts being listless) But if you perk up...(Acts perking up)
Hasa: Wake up?
T: (Acts perking up again) You see? The differences the way you are? In one, if you're not perky, you're like...(Acts being listless again) But when you perk up...(Acts perking up a third time) You see how big your eyes get?
Hasa: (Laughs)
T: Did I perk you up? You know, everybody looks like this...(Acts being listless a third time) And then everybody looks at me...(Acts perking up a fourth time) And-
Mar: Perk up.
T: You perk up.

In this example, the teacher is explaining the meaning of the phrase "perk up". While doing so, she acts out how it is like when one perks up by showing the change of the head, the eyes, the arms, and the torso. It was not coded as hand gesture but whole body acting because the movement of the arms is only part of the acting.

Head movements

In one of his books about gestures, McNeil (2005) points out that some head movements can be considered as the extension of gestures. For example, when the anatomical hand is immobilized or engaged, one might use the head as a third hand (e.g., pointing in a direction). In the current study, no such head movements were found. All the head movements are either the emblematic nodding or shaking. Below are two examples where the teachers nod or shake their head.

(44) Anh: I'm going to the bank. Can I say I will go to the bank?
T: I will go is ok. (Nods) The reason they choose I'm going to is because usually for plans we use going to. But it's ok, the grammar is ok if you say I will. (Nods)
Anh: (Nods)

(45) T: (Discussing paragraphs in book) However, the example provides hope that careful planning and creative thinking can lead to solutions to many of them. Ok, quick quiz, what is them?
Adu: Cities?
T: Cities? No. (Shakes head)
Adu: Eh people.
T: No, not people. (Shakes head)
Lyb: Example.
T: No.
Lyb: No? (Looks at book again)
Roma: Problem.
T: Problem. Right? Solutions to many of them, solutions to many of the problems.
Adu: Oh.
Lyb: Problem.
Others: (Write in book)

In the first episode, the teacher nods twice with additional verbal information to confirm the learner's language hypothesis. In the second example, the teacher shakes head twice when learners are trying to guess the antecedent of a pronoun.

Facial expressions

Sometimes teachers use facial expressions to confirm or disconfirm learners' utterances or opinions or to show that they have problems understanding what learners are saying because of a problematic utterance. For example:

(46) T: Now, give me some adjectives for Hasa (a student in the class).
Chen: Handsome.
Mar: Gorgeous.
Gra: Fun.
T: Fun, thank you, that's a good one.
Zhou: Stupid.
T: Zhou, how many times have I told you? (Shows helplessness on face)
Class: (Laughs)
Zhou: Oh, silly.
T: Yeah, silly is better.

(47) Cadi: Is there agency adoption? ("Agency" pronounced as /ei'd3ənsi/)
T: Uh? (Looks puzzled)
Cadi: Agency. (/ei'd3ənsi/)
T: Uh? (Looks puzzled)
Cadi: Agency. (/ei'd3ənsi/)
T: A--? (Looks puzzled and drags A sound)
Cadi: Agency. (/ei'd3ənsi/)
T: (Silent and still looks puzzled)
Cadi: A- (Tries to spell the word)
Lai: Agency. (/'eid3ənsi/)
T: Ok, the adoption agency.
Cadi: Yeah.

In the first example, the learner gives a pragmatically inappropriate word to describe a classmate. While giving verbal feedback, the teacher also shows helplessness on her face, which in a sense strengthens her disagreement and helps to prompt the learner to give the right word. In the second example, the teacher is not able to understand what the learner is saying because of a pronunciation problem. Four out of five teacher turns, he shows a puzzled look on his face, which, together with his short utterances, pushes the learner to repeat his utterance four times.

Combination

Like paralinguistic cues, teachers sometimes use more than one type of extralinguistic cues in the same episode. In this case, the multiple types of extralinguistic cues were coded as combination. For example:

(48) T: Do you know the meaning of desire?
Sar and others: No.
T: Desire means want. You want something.
Sar: (Indistinguishable)
T: Hmm?
Sar: (Indistinguishable)
T: Reversible?
Sar: Vegetable.
T: Vegetable?
Sar: Yeah.
T: No. (Shakes head) No, desire is something you want. (Points at heart) So I want, I want more fruit. That's the meaning.
Class including Sar: (No response)

In this example, the teacher uses two types of extralinguistic cues, head movement (shaking head) and gesture (pointing) when explaining the meaning of the word "desire". The type of extralinguistic cues in this episode is therefore combination.

The coding of gestures

Various coding systems for gestures have been proposed over the years. One of the most frequently adopted is the one established by McNeil (1992). In this scheme, there are four major categories: iconics, metaphorics, deictics, and beats. The current study also followed this coding tradition. In addition to the four major categories, the use of multiple types of gesture in the same episode was coded as a fifth category, combination, again for its potential to exert an influence on learners' noticing and learning different from a particular type of gesture by itself.

Iconics

Iconic gestures present images of concrete entities and/or actions. The form of the gesture and/or its manner of execution embodies the picturable aspects of semantic contents. Below is an example where the teacher uses iconic gestures to explain a word.

(49) Rus: What, what's a dumpster?
T: A dumpster is like a big, big, big box made of metal. Like this big. (Opens arms to show big) You dump trashes. (Hand gestures dumping) And a big machine comes and does this (hand gestures picking up things), wu-, wu-. Ok, and people were arrested... (Leaves no time for the learner to respond)

In this episode, the teacher uses three gestures as he explains the word "dumpster", a tangible object with a shape, color, and size in real life. Each gesture in turn represents a concrete thing or action, with the first one describing the size of a dumpster, the second one picturing the dumping action, and the third one depicting the picking up action.

Metaphorics

Metaphoric gestures present images of abstract ideas or concepts. One example McNeil (1992) gives is that the speaker appears to be holding an object as if presenting it but the meaning is not presenting an object but rather holding an idea or memory or some other abstract 'object'. Below is an example from the current study where a metaphoric gesture is used.

(50) Sal: (Looks puzzled) Might, may, more problem, more possibility...
T: May is a little more possibility. A little. (Hand gestures "a little" by putting two fingers close)
Sal: A little? What's more?
T: More, may.
Sal: May, more possibility.

In this episode, the teacher puts two fingers together as if two things were close to each other side by side. What she is really trying to convey is not the short distance between two objects but the idea of "a little". The concept expressed in the gesture is therefore abstract rather than concrete.

Deictics

Deictic gestures are pointing with the finger or other body parts. These gestures can refer to both the concrete and the abstract. In the current study, most deictic gestures fall into the former type. For example:

(51) Bia: Tell me the homework.
T: Just tell me the homework? Not if you want to be polite. Like I can say to you tell me the home work. (Points at self and then students)You can't say to me tell me the homework.
Bia: (Writes on book)

In this episode, the teacher points at herself and then her students as she explains the pragmatic aspects of the imperative. Both references, the teacher and the students, are concrete entities.

Beats

Beats are mere flicks of the hand(s) up and down or back and forth that seem to 'beat' time along with the rhythm of speech. Beat may look simple but the meaning behind it can be complex. It often signals the temporal locus in speech of something the speaker feels is important with respect to the larger discourse (McNeil, 1992). Below is an example where beats are used.

(52) Cadi: While he was studying at the ELC, he was playing basketball.

T: While he WAS studying, but he still IS. That's when we can't use the past because he IS still studying and he still IS playing basketball. That's present. (Stresses "was" and "is" and moves right hand up and down)
Cadi: (Nods)

In this example, the teacher is explaining why the learner should use the present tense "is" instead of the past tense "was" in his sentence. Every time he utters these words with a stress, he also moves his right hand up and down. The "verbal stress" from his mouth and the "nonverbal stress" of his hand work together to make these words stand out.

Combination

Sometimes teachers use more than one of the four types of gestures described above. In this case, the different types of gestures were coded together as combination. For example:

(53) Lyb: Excuse me, can you explain the second one? (Points at book)
T: (Looks at book) Ok, please pick it up and put it on. So it should go between pick and up (points at book) because that verb is a phrase. It's more than a word. So for some reason, I'm not sure why we do it this way, but whenever we have a phrasal verb and we're changing that noun to a pronoun we put the pronoun between the two parts of the phrasal verb. (Hand gestures "between" by raising and opening two fingers)
Lyb: (Writes in book)

In this episode, the teacher is explaining the position of the pronoun "it" in phrasal verbs. As she talks about it, she first points at the learner's book to show where it should go, and then opens two fingers to show its position. The first gesture is deictic and the second one is iconic. Now that two types of gestures are used, this episode was coded as having a combination of gestures.

3.6.1.3 The coding of teachers' metalanguage

James (1999) indirectly defines the concept of metalanguage when explaining metalanguage awareness. According to James, "meta" refers to the ability to use language to talk about language in the linguistic sense and cognition about cognition in the psychological sense. That is, linguistic metalanguage awareness means knowing what something is called, for example, part of speech and type of sociolects, while psychological metalanguage awareness means cognition about one's own knowledge, whether one can talk about it or not. In the current study, only metalanguage in the linguistic sense was examined.

Each episode was classified as being with or without metalanguage in the feedback. The metalanguage was then coded with a scheme adapted from Basturkmen et al. (2002). Two types of metalanguage were identified: technical terms and non-technical terms. Technical terms are those likely to be found in grammar books or linguistic references and often used by a small population such as teachers and linguists; non-technical terms are those that are commonly used by a larger population, including students (Basturkmen et al., 2002). Some of the examples for technical terms are "preposition", "irregular verb", and "yes/no question". Some of the examples for non-technical terms are "word", "mean", and "say". Like Basturkmen et al., if the same term occurred multiple times in the feedback of the same episode, it was counted only once. However, when different terms were used in the feedback of the same episode, the occurrence of each term was counted once. The following example illustrates how the metalanguage in teacher feedback was coded:

(54) Sal: What's the difference between be and to be?
T: To be is the *infinitive*. Be is the *base form*, *base form* be.
Sal: (Writes on paper)

In the teacher's feedback, there are two different technical terms which can be found in linguistic references: "infinitive" and "base form". Although they occur in the same teacher turn, they were both counted. On the other hand, the term "base form" was only counted once even though it occurs twice in the same teacher turn. Ultimately, there are two occurrences of teacher metalanguage in this episode.

After all metalanguistic terms were labeled, the feedback episodes were further analyzed into three categories according to the properties of the terms: those with only technical terms, those with only non-technical terms, and those with both technical and non-technical terms. For example:

(55) Ye: I were-
T: I WAS, were is *plural*. (Stresses "was")
Ye: Yeah, I was doing my homework.

(56) Lo: White and black. Write them white and black.
T: We *say* black and white.
Lo: Oh ok. Black and white.

(57) Wa: Riot, I don't know how to pronounce riot. Is it riot? (/ræt/)
T: What's the *word*? Ri--ot, ri--ot. Two *syllables*. (Drags voice between /rai/ and /ət/)
Wa: Riot.
T: But people *say* it quickly, riot.
Wa: Riot, riot.

In the first episode, the teacher uses the term "plural", a technical term which can be found in most grammar glossaries. In the second episode, the teacher uses the word "say", a common word which occurs very often in daily conversations. In the third episode, the teacher uses two terms: "word" and "syllable", with "word" as a non-technical term and

"syllable" as a technical term. Teachers' metalanguage in the three episodes was therefore categorized as technical term only, non-technical term only and technical plus non-technical term.

As Basturkmen et al. (2002) note, "it is realised that the distinction between technical and non-technical terms is not always straightforward and some metalinguistic terms fall somewhere between the two ends of the continuum" (p.5). To avoid over-subjectivity, I consulted three grammar books: *Portable English Handbook* (Herman, 1982), *Introducing English Grammar* (Leech, 1992), and *A Glossary of English Grammar* (Leech, 2006). Now that the metalanguage examined in the current study is not just grammar terms, I also consulted a native speaker and an ESL speaker about non-grammar terms that fall between the technical and non-technical continuum. Both parties provided their insights on the property of these terms according to their "feel" and daily language experience. It was hoped that the knowledge and perspectives of "outsiders" would help objectify my decision-making. When a term was still difficult to decide after this, I would discuss it with another SLA researcher.

3.6.1.4 The coding of the occurrence and successfulness of uptake

In coding learner uptake, a scheme was again adapted from Loewen (2004) as shown in Table 3. In the first step, depending on whether learners responded to teacher feedback and whether they had the chance to do so, three categories were identified: uptake, no uptake, and no opportunity. Any response to teachers' feedback, even ambiguous expressions such as "Oh" or "Yeah", was coded as uptake. Otherwise they were coded as

no uptake or no opportunity. The difference between no uptake and no opportunity is that in no uptake learners have the opportunity to respond but choose not to do so, while in no opportunity learners have no chance to respond because of factors such as topic continuation. It is important to note that now that a video camera was used in the current study, the recordings show that many times learners responded to teachers' feedback by acting, nodding, or writing with or without linguistic utterances. Such reactions were also learners' responses to teachers' feedback. Acting, nodding and writing, even when there was no verbal utterance accompanying them, were therefore counted as uptake.

Table 3 Coding scheme for learner uptake

	Definition	Categories
Occurrence of uptake	Whether student responds to feedback	*Uptake*: student responds to feedback *No uptake*: student has the chance to respond to feedback but does not respond *No opportunity*: student does not have the chance to respond to feedback
Successfulness of uptake	How the quality of student response to feedback is	*Successful uptake*: student incorporates correct form in production *Partially successful uptake*: Student incorporates some of the correct form in production but not all of it so that production is still not completely correct. *Unsuccessful uptake*: student tries but fails to incorporate correct form in production *Acknowledge*: Student simply acknowledges feedback with utterances such as oh and year or by nodding. *Inconclusive*: unclear whether student successfully incorporates the correct form in production

Adapted from Loewen (2004, p.166)

The second step is the coding of the successfulness of uptake. When learners incorporated the correct form in their production, the response was coded as successful uptake. When learners tried to incorporate the correct form but only did part of it, the response was coded as partially successful uptake. When learners tried but totally failed

to incorporate the correct form, it was coded as unsuccessful uptake. When learners reacted to teachers' feedback by simply saying "Yeah", "Oh", etc, the response was coded as acknowledge. Acknowledge was not operationalized as unsuccessful uptake because although it is possible that learners did not understand the feedback but simply said something to indicate that they were listening, it is also possible that they might have indeed understood the correct form but decided not to take the trouble to incorporate it in a new language production. A fifth category was inconclusive. In inconclusive uptake, it is difficult to determine if the correct form was produced or understood due to the quality of the recordings, choral responses, etc. For example:

(58) T: Eh unadorn.
Gra: Yes
Hasa: No.
T: What is that? Gra, you said yes, what's unadorn?
Gra: Eh... (Chuckles)
T: Ok. Students, you know, you know, what are these? (Touches earrings) Mik?
Hasa: Earrings.
T: Mik?
Mik: I don't know.
T: That's why I'm trying to help you understand. What're these?
Class: Earrings.
T: Why do you believe I put the earrings on? Why?
Hasa: To look, to look beautiful.
T: To adorn. Ok?
Class: Oh.
T: To make me look nice. Now if I take them off, then I would be doing what?
Class: (Indistinguishable)
T: Then I would be what?
Mar: Gorgeous. (Laughs)
T: Unadorn.
Hasa: No, no.
T: When you adorn, it's that you put things on to make your, to make yourself look nice. Ok? But when you take them off, then it's [unadorn]. (Shows putting on and taking off earrings)
Class: [Unadorn].

In this example, there are a few points where more than one learner responds to the teacher at the same time and it is impossible to distinguish who are speaking. Hasa indicates that he does not understand the word "unadorn". Even though he participates in the discussion all the way toward the end of the episode, it is unclear from neither the audio tape nor the video tape if he is one of those students who finally say the word together with the teacher.

The examples in Table 4 illustrate how the occurrence and the successfulness of uptake were coded. In Example 59, the learner has the opportunity to respond to the teacher's feedback but he chooses to talk about something else. In Example 60, the learner does not get the chance to respond to the teacher's recast as a result of the teacher's topic continuation. The uptake move for these two episodes was therefore coded as no uptake and no opportunity respectively. In the next four examples, learners do respond to teachers' feedback, but in different ways. In Example 61, the learner successfully incorporates the correct form in his production. In Example 62, the learner is able to give a partially correct answer at the teacher's elicitation but still leaves out the object of the word "move" after the teacher gives him the correct form. His final production therefore is still grammatically flawed. In Example 63, the learner fails to produce the correct form after a few tries. In Example 64, the learner responds to the teacher's feedback with a simple "ah" and it is not clear if he really understands it. Based on the different kinds of response moves by learners, the successfulness of uptake in these episodes was coded as successful, partially successful, unsuccessful, and acknowledge respectively.

Table 4 Examples for the coding of the occurrence and successfulness of uptake

Episodes	Occurrence of uptake	Successfulness of uptake
(59) T: Think about your day. When do you have ten or fifteen minutes when you're not necessarily doing something that you can work on vocabulary? Saud: In the bus. T: On the bus. Saud: Waiting before the class.	No uptake	--
(60) Rus: Grinding to a halt. (/'grindiŋ/) T: Ok, grinding to a halt. Grinding. How many people here know what grinding means?	No opportunity	--
(61) Squ: One people speak one people write? T: One PERSON can speak. (Stresses "person") Squ: Oh one person. I'm sorry.	Uptake	Successful
(62) Mato: The answer is let's go? (Could you please help me move the sofa?) T: Let's go is like let's go away outside. (Hand gestures going away) Let's- (Stops) Mato: Move. T: Move it. Yeah. Mato: Let's move.	Uptake	Partially successful
(63) Ne: I would some coffee. T: I would some coffee? (Rising intonation) You need a verb. I would like. Ne: I would s-, I would some... T: (Shakes head) When you use, when you use would, would is like a helping verb, so you have to have a main verb. Ne: I would some coffee? T: (Shakes head and shakes hand) Would is not by itself. It's not alone. Ne: (Nods) But I...some kind of...	Uptake	Unsuccessful
(64) T: So what kind of strategies have we used so far today? Stor: Strategy? T: Strategies means a way of doing something, a plan for how to do it. Stor: Ah. (Writes in notebook)	Uptake	Acknowledge

In the coding of the successfulness of uptake, one thing that deserves a special note is again students' nonverbal responses to teacher feedback, including acting, nodding and writing. On some occasions, learners' nonverbal responses clearly indicate that they have understood the feedback. In this case, the uptake was coded as successful. For example,

(65) T: How about beady? Beady?
Class: (Silence, 4 seconds)
T: Beady? Ok guys, watch, Tona, watch. (Acts) Look at my eyes. Now if I put them...(Acts) So when I, when I put them like this and I go...(Acts)
Class: (Laugh)
Sa: (Imitates)
T: Sa! But that's beady. See the eyes? They get like a little like this like, but yeah the eyes they're going like this like looking all over the place. (Acts) Beady, ok? Beady, beady.
Class: Beady, beady eyes.

In this episode, even though it is not clear if Sa is among the students who repeat the word "beady" in the end, her imitation of beady eyes suggests that she understands what the teacher is saying. The uptake move for Sa in this episode was therefore coded as successful instead of inconclusive.

On other occasions, learners responded to teacher feedback by nodding. This might happen with or without verbal utterances. When there was a verbal utterance, the successfulness of the uptake was decided by the verbal utterance. When there was no verbal utterance, the nodding was coded as acknowledge. For example:

(66) Xin: Pot of money means a big amount of money?
T: Sorry?
Xin: Pot of money.
T: Pot of money? Where do you see that?
Xin: (Shows T word in book) Pot of money.
Ira: Pot of money.
Xin: Refers to big sum of money?
T: No, not necessarily a lot. I think she actually means a physical pot. (Hand gestures "pot") She has some sort of bowl or pot or jar that she puts money into.

But it doesn't mean something necessarily physical. It could also mean just, just a collection of money. (Hand gestures "collection") When you play poker, the money you put in the middle when you bet, that's the pot. The winner takes the money. It's not really a pot. (Hand gestures "pot") It's just a pile of that money. (Hand gestures "pile")
Xin: (Nods)
T: Everyone just puts money into the pot. We call that a pot. Yeah.
Xin: (Nods)

(67) Bibo: ATM card, you can eh take your money in ATM machine?
T: Take it FROM the ATM machine. (Stresses "from")
Bibo: Yeah yeah yeah. (Nods)

(68) Lyb: Yes, the farm now belongs the animal.
T: Belongs TO the animal, right? (Stresses "to") Belongs to the animal.
Lyb: Belongs to the animal. (Nods)

In Example 66, Xin responds to the teacher's explanation of the phrase "pot of money" with nodding without saying a word. The uptake in this episode was therefore coded as acknowledge. In the other two examples, nodding is both accompanied by an utterance. In Example 67, the utterance is the typical acknowledgement token "yeah" while in Example 68, the student correctly incorporates the teacher's feedback in her new utterance. The uptake in these two episodes was therefore coded as acknowledge and successful uptake respectively.

The coding of students' writing move with verbal utterances is similar to that of nodding with verbal utterances, but writing without verbal utterances is different from nodding without verbal utterances. When students start to write after a feedback move, it is not clear whether they are writing what the teacher is saying or something that is not related to the particular episode. Writing alone was therefore coded as inconclusive uptake. For example:

(69) Rus: What do you call this? (Points at picture)

T: Ok, I'll say it exactly. It's called a riot baton or riot stick. Riot stick or riot baton or a night stick. They hit people with a night stick. (Hand gestures hitting)
Rus: (Similes and writes on paper)

In this episode, the student writes on a piece of paper after the teacher's feedback. Even though it is quite possible that she is writing down riot baton or night stick, there is no observable evidence that this is indeed what she is writing. The uptake in this episode was therefore coded as inconclusive.

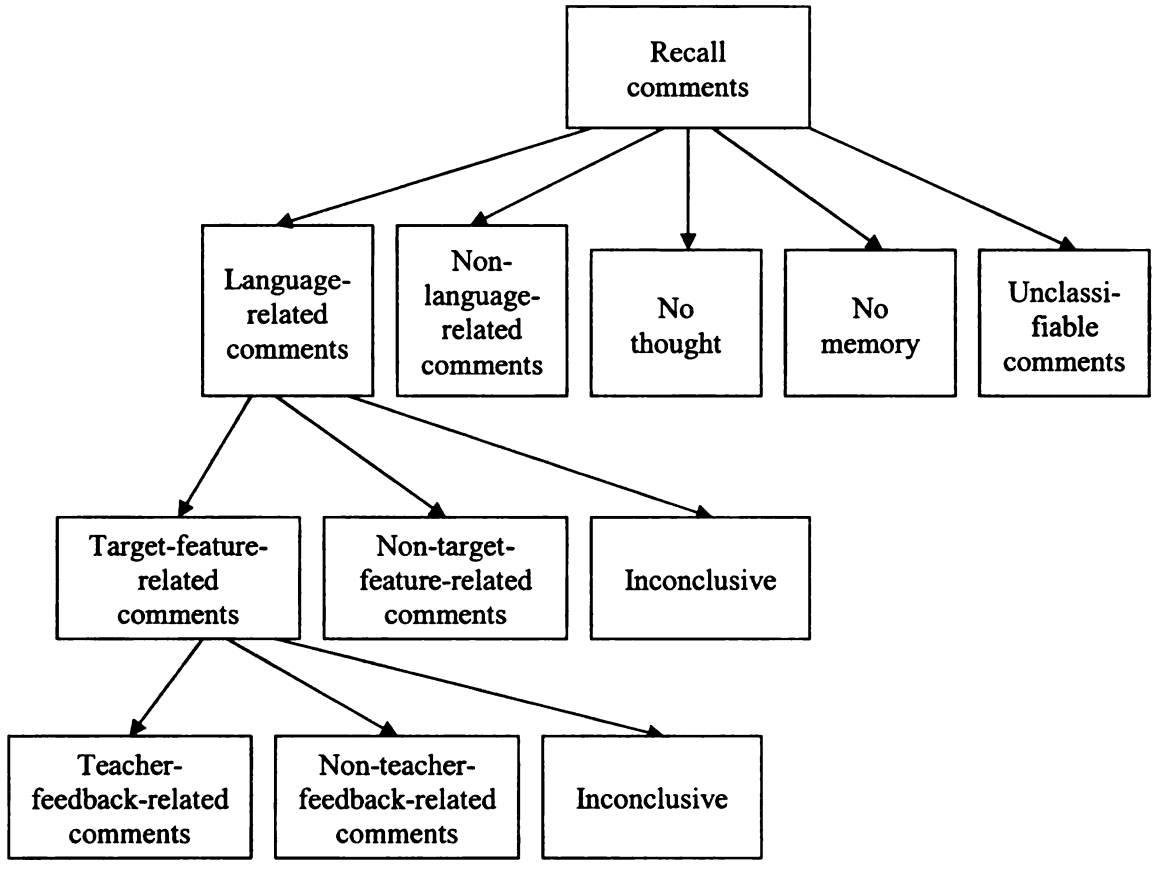
3.6.2 The coding of stimulated recall comments

The analysis process of the comments made by student participants in the stimulated recall interviews is shown in Figure 4. First, the overall recall comments were classified into five categories: language-related comments, non-language-related comments, no thought, no memory, and unclassifiable comments. Next, language-related comments were further classified into target-feature-related, non-target-feature-related, and inconclusive comments. Finally, target-feature-related comments were put into three categories: teacher-feedback-related, non-teacher-feedback-related, and inconclusive comments.

3.6.2.1 The coding of overall recall comments

In coding the overall recall comments, a scheme was adapted from Egi (2004). The comments were classified into five categories: language-related comments, non-language-related comments, no thought, no memory, and unclassifiable comments.

Figure 4 Coding process of stimulated recall comments



Language-related comments

This category was operationalized as any comment that involves linguistic items or language structures with or without participants' reference to their own error production or the teacher's feedback. The comments range from simple observational remarks to those involving linguistic analyses. For example:

(70) I forgot what's that. The teacher remembered to us imperative.

(71) I just concentrate on the two words and I found the differences that when we say pay BACK, it's not a compound. There's a blank between it. So I just considered it as a verb and preposition. When we don't link those two together, it's probably the verb. But payback is actually one word. So despite of the verb, it should be the noun. When it's a compound, it's a noun.

In these two examples, the first comment is a simple observation of what the participant heard. In the second comment, the learner made a detailed analysis on the spelling, pronunciation, structure, and part of speech of "pay back" and "payback".

Non-language-related comments

In this category, participants talked about topics that are not related to language issues, for example, the task at hand, other learners' behavior in class, teachers' teaching method, etc. For example:

(72) I was thinking I disagree with the article in the book.

(73) I think Kuwa makes a lot of questions. I said, ok Kuwa, you make a lot of questions. Some are boring. Some were not that helpful for me.

(74) She's only speaking to one, two, three, four, four people. She must speak to all these people. But maybe other people understand. Maybe she asked other people already. I feel bad if the teacher didn't explain to everybody. I took a course in education in my country. I know some education. Yeah, that's I was thinking.

In these three examples, the first one is an opinion statement about a question the class was discussing; the second one is a remark about another learner's action in class; and the third one is a comment on the way the teacher was talking to students. All three comments are somewhat related to the lesson but none is related to language issues.

No memory

This category refers to participants' report that they do not remember their thoughts regarding a particular feedback episode. For example, one learner said:

(75) I don't remember. Sorry.

No thought

This category refers to participants' report that they had no specific thoughts during a particular feedback episode. For example, a learner said:

(76) I don't think nothing because I was bored. That's enough.

Unclassifiable comments

This category was operationalized as any comment that cannot be classified into any of the categories above. For example:

(77) Any thought? It's very interesting. I didn't so relax...(inaudible) because me and Rin, especially the student sitting at the lastest place in the class.

In this example, it is hard to understand what the learner was saying and whether he was talking about the lesson or something else. This comment was therefore coded as unclassifiable.

3.6.2.2 The coding of language-related comments

After language-related comments were identified, they were further categorized into target feature-related, non-target-feature-related, and inconclusive comments.

Target-feature-related comments

This category was operationalized as any comment that is related to the language feature which had been targeted at by the teacher in a feedback episode, with or without the learner' reference to the feedback. For example:

(78) Yes, I can remember. She is explain two words, what's the difference between the words. The first word is "give" and the second word is "taking". She said many

students is confused, so she explain with example. I think this is a very good example, very simple example.

In this comment, the learner was talking about the difference between "imply" and "infer", two words the teacher had explained in detail in a feedback episode, and how the teacher explained it. It was therefore coded as target-feature-related.

Non-target-feature-related comments

This category was operationalized as comments that are related to a language feature which was not the target of the teacher's feedback or comments where the learner was talking about the same language item as the teacher but with a different linguistic focus. For example:

(79) About the word. I didn't know this word at all but I wanted to get it from word root, prefix or suffix. With S-O-L, I thought of "solid". Then I thought of "solo". "Solo" means alone. So I thought this word might be related to that. Then "tude", I don't know why I associated with "attitude". So I thought "solitude" might mean an attitude or idea.

In the original feedback episode, the teacher was giving feedback about the spelling of the word "solitude", but in the comment the learner was talking about the meaning of the word. Since the learner's focus is different from that of the teacher, it was coded as non-target-feature-related.

Inconclusive

This category was operationalized as comments that do not fall into either of the two categories above. For example:

(80) He said something from. I told him this is not polite. I don't know this meaning. Don't say your name. (Draws on paper) Don't say from, just put something. He told me to use from only when I write friend or family.

This comment comprises six separate sentences which do not seem to be closely connected. It can be seen that the student was talking about language use, but it is very difficult to tell whether he was talking about the word "sincerely", which was the target of the original feedback episode.

3.6.2.3 The coding of target-feature-related comments

Finally, target-feature-related comments were put into three categories: teacher-feedback-related, non-teacher-feedback-related, and inconclusive.

Teacher-feedback-related comments

This category was operationalized as any comment where students referred to teacher feedback or things that they thought of as the result of the feedback. For example:

(81) I didn't know about that word. Me and my partner asked this question about this word. And then she explained more specifically, and at that time she gave us example, so I could understand more easily.

(82) When I heard of the word (classic), I thought of classic music. And then he said classic coke. That reminded me of the NCA I'd been watching. Coca cola did a lot of commercials there. So I was thinking of the coca cola commercials.

In the first comment, the learner explicitly referred to the teacher's use of examples and how that helped with his understanding of a word. In the second comment, although the learner got "carried away" to talk about the coca cola commercials instead of the meaning of the word "classic", which the teacher was explaining in the original episode, the thought of the coca cola commercials is a result of hearing the teacher's use of the classic coke example. It was therefore still coded as teacher-feedback-related.

Non-teacher-feedback-related comments

In contrast to teacher-feedback-related comments, non-teacher-feedback-related comments are those without learners' reference to teacher feedback or things they thought of as the result of the feedback. For example:

(83) Yesterday I was thinking, at that time, I was thinking about the spelling of this word because we was, we were, me and my partner, we were struggling and have a little problem how to write this word, solitude. That's it. Yeah, I was thinking how to write the spelling of the word.

In this example, the learner didn't say anything about the teacher's feedback about the spelling of the word "solitude" but how he himself was trying to figure it out. It was therefore coded as non-teacher-feedback-related.

Inconclusive

Similar to comments which are neither target-feature-related nor non-target-feature-related, this category was operationalized as comments that are neither teacher-feedback-related nor non-teacher-feedback-related. For example:

(84) It's the same word in Spanish, motivation. I was thinking, that's the same word in Spanish!

In the original episode, the teacher and the class were talking about the meaning of the word "motivation". In the comment, the learner was talking about the similarity between the English word and the Spanish word. It is not clear if she was talking about the meaning or the pronunciation of the word or both. Nor is it clear if the student's thought is a result of her hearing the sound of the word or the teacher's explanation. It was therefore classified as inconclusive.

3.6.3 The coding of test results

The results from the immediate and delayed test were analyzed with Loewen and Philp's (2006) coding scheme. Learners' answers to the test items were sorted into three categories: correct, incorrect, and partially correct. When the answer a learner provided matched the correct form that had been targeted at in a feedback episode, it was rated as correct. When a learner failed to provide the correct form that had been targeted at, the answer was rated as incorrect. When a learner's answer showed some improvement toward the target form but still needed repair, it was rated as partially correct. The examples in Table 5 show how the test results were coded.

Table 5 Examples for the coding of test results

Episodes	Prompts	Responses	Ratings
(85) Lyb: My husband studied in Florida. T: Oh really? Lyb: Yeah. Before ten years. T: He studied there ten years AGO. (Stresses "ago") Lyb: Yes. T: Not before ten years. Ten years AGO. (Stresses "ago") Lyb: Ah ten years ago, yes yes.	The following sentence is inappropriate. Please listen carefully and tell me how you could make it better. My husband studied in Florida before ten years.	My husband studied in Florida ten years ago.	Correct
(86) T: What is this? What's this thing? Rouf: Skull. (/skul/) T: A skull. Ok. Do we have a skull? (/skʌl/). Class: Yes.	Please read the following sentence and word. Everyone has a skull. Skull	The word "Skull" still pronounced as /skul/	Incorrect
(87) Mato: It is going to not to be to look good. T: It's going to look terrible. Mato: Terrible. T: It is going to look terrible. Mato: Terrible.	The following sentence is inappropriate. Please listen carefully and tell me how you could make it better. The new building is going to not to be to look good.	The new building is going to be to look terrible.	Partially correct

In the first example, the learner successfully corrected an ungrammatical sentence in the way the teacher did in the feedback episode. In the second example, the learner pronounced the word "skull" as /skul/ again as he did during the classroom interaction. In the last example, the teacher made three changes to the learner's original sentence: deleting the negation word "not", deleting the "be" verb, and changing the adjective "good" into "terrible". In the test, the learner deleted "not" and changed "good" into "terrible" but kept the "be" verb, thus producing a sentence that is better but still ungrammatical. The rating for the answers to the three questions is therefore correct, incorrect, and partially correct respectively.

3.6.4 The reliability of coding

In order to increase the reliability of coding, a second rater coded 11.2% of the feedback episodes according to the seven characteristics, nonlinguistic cues and metalinguistic terms in teacher feedback, and the occurrence and successfulness of uptake. 10.8% of the stimulated recall comments and 11.9% of the test results were also coded by a second rater. The percentage of agreement for each variable is shown in the Table 6. It ranges from 78.1% for the source of feedback episodes to 98.3% for test results.

Table 6 Reliability of coding

Variables	Percentage of agreement
Type	84.4%
Source	78.1%
Response	81.9%
Linguistic focus	88.8%
Directness	90.0%
Complexity	90.0%
Emphasis	93.1%
General paralinguistic cues	97.5%
Type of paralinguistic cues	88.0%
General extralinguistic cues	98.1%
Type of extralinguistic cues	93.7%
Type of gestures	83.3%
General metalanguage	86.9%
Type of metalanguage	89.4%
Occurrence of uptake	91.3%
Successfulness of uptake	87.1%
Recall comments	93.3%
Test results	98.3%

3.6.5 Statistical analysis

After the coding of all data, Pearson's chi-square tests were performed with SPSS 12.0 to see if there is a significant relationship between the dependent variables (the occurrence of uptake, the successfulness of uptake, test results, and stimulated recall comments) and the independent variables (the characteristics of feedback episodes, the nonlinguistic cues in teacher feedback, and the metalanguage in teacher feedback). The conventional alpha value of .05 was selected as the cutting point. In order to meet the requirement of chi-square tests that no more than 20% of expected cell counts should be less than 5, a trial test was performed for each pair of dependent and independent variables. According to

the results of the trial tests, categories with small counts were either excluded or conflated with other categories as necessary and appropriate. Table 7 shows the changes of the categories within variables. All changes are noted again in or at the bottom of corresponding tables except three, namely, the combination of partially successful and unsuccessful uptake, the combination of inconclusive teacher-feedback-related comments and the category of "other", and the combination of partially correct and incorrect test results. These three were highly frequent ones. It would be repetitive if they were stated again and again.

Table 7 Treatment of variables in chi-square analysis

Variables	Categories	Treatment
Type	Teacher-initiated, student-initiated, reactive	No change
Source	Code, message, semantic	Semantic and message combined when dependent variable is delayed test results and recall comments
Response	Provide, elicit	No change
Linguistic focus	Vocabulary, grammar, pronunciation, spelling	Spelling excluded when dependent variable is immediate test results and recall comments
Directness	Direct, indirect	No change
Complexity	Simple, complex	No change
Emphasis	Light, heavy	No change
(General) paralinguistic cues	Present, absent	No change
Type of paralinguistic cues	Word stress, rising intonation, eliciting stop, dragging voice, mimicking sound, combination	Mimicking sound excluded from all chi-square tests; combination excluded when dependent variable is the occurrence and successfulness of uptake; rising intonation excluded when dependent variable is immediate test results

Table 7 (cont'd)

(General) extralinguistic cues	Present, absent	No change
Type of extralinguistic cues	Gestures, whole body acting, head movements, facial expressions, combination	Facial expressions excluded when dependent variable is successfulness of uptake and test results; Whole body acting excluded when dependent variable is test results
Gestures	Iconics, metaphorics, deictics, beats, combination	Combination excluded when dependent variable is immediate test results; beats and combination excluded when dependent variable is delayed test results
(General) metalanguage	Present, absent	No change
Type of metalanguage	Tech-only, non-tech-only, tech+non-tech	No change
Occurrence of uptake	Uptake, no uptake, no opportunity	No change
Successfulness of uptake	Successful, partially successful, unsuccessful, acknowledge, inconclusive	Partially successful and unsuccessful combined; both excluded when dependent variable is test results and when independent variable is type of paralinguistic cues and type of extralinguistic cues
Recall comments	Teacher-feedback-related, non-teacher-feedback-related, inconclusive, other	Inconclusive combined with "other" in all chi-square tests
Test results	Correct, incorrect, partially correct	Partially correct and incorrect combined in all chi-square tests

While a chi-square value can tell whether the relationship between two variables is statistically significant or not, it does not specify how much each category contributes to the rejection of the null hypothesis. In order to get this information, adjusted standardized residuals (expressed in the present study as ASR) were computed along with chi-squares. The conventional 2.0 in absolute value was selected as the cutting point. Now that adjusted standardized residuals were used to find the major contributors of significant chi-square values, they were not examined when a chi-square value is greater than .05

(See Hinkle et al., 2003). To economize language and save space, the term "adjusted standardized residuals" is presented as "residuals" in tables and table titles. For the same reason, when a variable (dependent or independent) only has two categories in it and when the chi-square is 2x2, only the adjusted standardized residuals of one category are presented and noted now that the residuals of the other category always have the same absolute values but only with the opposite property.

3.7 Summary

In summary, this chapter addressed methodological issues such as instruments and procedures, detailed the coding schemes for the observation data, learners' stimulated recall comments, and test results, and reported the reliability of coding. In addition, it briefly explained the statistical analysis of the whole data set. The next three chapters report and discuss the findings from the analysis.

CHAPTER 4

CHARACTERISTICS OF FEEDBACK EPISODES, THE NOTICING AND THE EFFECT OF TEACHER FEEDBACK

4.1 Introduction

This chapter addresses the first research question: Do the characteristics of teacher feedback episodes affect learners' noticing and learning? After presenting the distribution of feedback episodes according to their general characteristics, I report and discuss the relationship between these characteristics and learners' responses to teachers' feedback, learners' comments in the stimulated recall interviews, and results from both the immediate test and delayed test.

4.2 Distribution of feedback episodes by characteristics

This section reports the distribution of feedback episodes according to their general characteristics: type, source, response, linguistic focus, directness, complexity, and emphasis. As shown in Table 8, in terms of type, there is no big difference among the three kinds of feedback episodes, with reactive episodes (35.1%) slightly more than student-initiated episodes (31.4%) and teacher-initiated episodes (33.5%). In terms of source, both code-related episodes (59.8%) and message-related episodes (36.8%) occurred noticeably more often than semantic episodes (3.4%) and code-related episodes noticeably more often than message-related episodes. In terms of teachers' response type, the percentage of provide is as high as 92.1% while that of elicit is as low as 7.9%, with an 84.2% difference between the two. In terms of linguistic focus, vocabulary comes first

Table 8 Distribution of feedback episodes by characteristics (N=1434)

Characteristics	Number	Percentage
Type		
T-initiated	481	33.5%
S-initiated	450	31.4%
Reactive	503	35.1%
Source		
Code	858	59.8%
Message	527	36.8%
Semantic	49	3.4%
Response		
Provide	1321	92.1%
Elicit	113	7.9%
Linguistic focus		
Vocabulary	712	49.7%
Grammar	532	37.1%
Pronunciation	152	10.6%
Spelling	38	2.6%
Directness		
Direct	1107	77.2%
Indirect	327	22.8%
Complexity		
Simple	572	39.9%
Complex	862	60.1%
Emphasis		
Light	182	12.7%
Heavy	1252	87.3%

with a percentage of 49.7%; grammar comes next with a percentage of 37.1%; pronunciation comes third with a percentage of 10.6%; and spelling comes last with a percentage as low as 2.6%. In terms of directness, the majority of feedback episodes are direct (77.2%) rather than indirect (22.8%). In terms of complexity, 39.9% of the

episodes are simple and 60.1% are complex. Last, in terms of emphasis, the percentage for heavy episodes is 87.3% while that of light episodes is only 12.7%.

To sum up, reactive, code-related, direct, complex, and heavy episodes occurred more often than teacher/student-initiated, message-related/semantic, indirect, simple, and light episodes in the classrooms observed in this study. In most of the feedback episodes, teachers provided direct information instead of eliciting answers from learners. The linguistic focus of teachers' feedback is mainly vocabulary and grammar.

4.3 Characteristics of feedback episodes and learner uptake

This section reports the relationship between the general characteristics of feedback episodes and learner uptake. After an overview of the occurrence and successfulness of uptake, the two variables are presented together according to each characteristic in the order of type, source, response, linguistic focus, directness, complexity, and emphasis.

Overview of the occurrence and successfulness of uptake

Table 9 shows the occurrence and successfulness of uptake. In 65.2% of a total of 1434 episodes, learners responded to teachers' feedback in some way. In 21.3% of the episodes, they did not produce any uptake even though there was a chance for them to do so. In 13.5% of the episodes, they did not have the opportunity to respond. When uptake did occur, 45.2% was successful; 0.6% was partially successful; 1.2% was unsuccessful, 37.6% fell into the acknowledge category, and 15.3% was inconclusive.

Table 9 Overview of the occurrence and successfulness of uptake

	Number	Percentage
Occurrence of uptake		
Uptake	935	65.2%
No uptake	306	21.3%
No opportunity	193	13.5%
Total	1434	100%
Successfulness of uptake		
Successful	423	45.2%
Partially successful	6	0.6%
Unsuccessful	11	1.2%
Acknowledge	352	37.6%
Inconclusive	143	15.3%
Total	935	100%

Type

Tables 10, 11, 12, and 13 illustrate the occurrence and successfulness of learner uptake according to the type of feedback episodes. Table 10 shows that learners responded to teacher feedback in 54.7% of teacher-initiated episodes, 71.6% of student-initiated episodes, and 69.6% of reactive episodes. Adjusted standardized residuals in Table 11 reveal that uptake was significantly more frequent in student-initiated and reactive episodes while no uptake and no opportunity were significantly more frequent in teacher-initiated episodes. Table 12 shows that when uptake did occur, 45.2% was successful in teacher-initiated episodes, 30.4% was successful in student-initiated episodes, and 58.9% was successful in reactive episodes. Adjusted standardized residuals in Table 13 reveal that reactive episodes led to significantly more successful uptake, student-initiated episodes led to significantly more acknowledge, and teacher-initiated episodes led to significantly more inconclusive uptake. Chi-square tests indicate that both the occurrence

and successfulness of uptake are significantly related to the type of feedback episodes (p=.000 in both cases).

Table 10 Type and the occurrence of uptake

	Number of episodes	Occurrence of uptake		
		Uptake	No uptake	No opportunity
T-initiated	481	263 (54.7%)	122 (25.4%)	96 (20.0%)
S-initiated	450	322 (71.6%)	97 (21.6%)	31 (6.9%)
Reactive	503	350 (69.6%)	87 (17.3%)	66 (13.1%)

X^2(4, n=1434)=49.523, p=.000

Table 11 Type and the occurrence of uptake residuals

	Uptake	No uptake	No opportunity
T-initiated	-5.9	2.6	5.1
S-initiated	3.4	.1	-4.9
Reactive	2.6	-2.7	-.3

Table 12 Type and the successfulness of uptake

	Number of uptake	Successfulness of uptake			
		Successful	Unsuccessful	Acknowledge	Inconclusive
T-initiated	263	119 (45.2%)	3 (1.1%)	89 (33.8%)	52 (19.8%)
S-initiated	322	98 (30.4%)	4 (1.2%)	167 (51.9%)	53 (16.5%)
Reactive	350	206 (58.9%)	10 (2.9%)	96 (27.4%)	38 (10.9%)

X^2(6, n=935)=69.521, p=.000

Table 13 Type and the successfulness of uptake residuals

	Successful	Unsuccessful	Acknowledge	Inconclusive
T-initiated	.0	-1.0	-1.5	2.4
S-initiated	-6.6	-1.0	6.5	.7
Reactive	6.5	1.8	-5.0	-2.9

Source

Tables 14, 15, 16, and 17 illustrate the occurrence and successfulness of learner uptake according to the source of feedback episodes. Table 14 shows that learners responded to

teacher feedback in 69.8% of code-related episodes, 56.4% of message-related episodes, and 79.6% of semantic episodes. Adjusted standardized residuals in Table 15 reveal that uptake was significantly more frequent in code-related and semantic episodes while no uptake and no opportunity were significantly more frequent in message-related episodes. Table 16 shows that when uptake did occur, 51.8% was successful in code-related episodes, 33.7% was successful in message-related episodes, and 33.3% was successful in semantic episodes. Adjusted standardized residuals in Table 17 reveal that code-related episodes led to significantly more successful uptake while message-related and semantic episodes led to significantly more acknowledge. There is no significant difference among the three in terms of unsuccessful and inconclusive uptake. Chi-square tests indicate that both the occurrence and successfulness of uptake are significantly related to the source of feedback episodes (p=.000 in both cases).

Table 14 Source and the occurrence of uptake

	Number of episodes	Occurrence of uptake		
		Uptake	No uptake	No opportunity
Code	858	599 (69.8%)	169 (19.7%)	90 (10.5%)
Message	527	297 (56.4%)	132 (25.0%)	98 (18.6%)
Semantic	49	39 (79.6%)	5 (10.2%)	5 (10.2%)

$X^2(4, n=1434)=34.345, p=.000$

Table 15 Source and the occurrence of uptake residuals

	Uptake	No uptake	No opportunity
Code	4.5	-1.9	-4.0
Message	-5.4	2.6	4.3
Semantic	2.2	-1.9	-.7

Table 16 Source and the successfulness of uptake

	Number of uptake	Successfulness of uptake			
		Successful	Unsuccessful	Acknowledge	Inconclusive
Code	599	310 (51.8%)	13 (2.1%)	175 (29.2%)	101(16.9%)
Message	297	100 (33.7%)	3 (1.0%)	155 (52.2%)	39 (13.1%)
Semantic	39	13 (33.3%)	1 (2.6%)	22 (56.4%)	3 (7.7%)

$X^2(6, n=935)=52.206, p=.000$

Table 17 Source and the successfulness of uptake residuals

	Successful	Unsuccessful	Acknowledge	Inconclusive
Code	5.3	1.1	-7.1	1.8
Message	-4.8	-1.3	6.3	-1.3
Semantic	-1.5	.4	2.5	-1.3

Response

Tables 18 and 19 illustrate the occurrence and successfulness of learner uptake according to the response type of feedback episodes. Table 18 shows that learners responded to teacher feedback in 62.4% of provide episodes and 98.2% of elicit episodes. Adjusted standardized residuals in the table reveal that uptake was significantly more frequent in elicit episodes while no uptake and no opportunity were significantly more frequent in provide episodes. Table 19 shows that when uptake did occur, 38.7% was successful in provide episodes and 93.7% was successful in elicit episodes. Adjusted standardized residuals in the table reveal that elicit episodes led to significantly more successful uptake while provide episodes led to significantly more acknowledge and inconclusive uptake. There is no significant difference between the two in terms of unsuccessful uptake. Chi-square tests indicate that both the occurrence and successfulness of uptake are significantly related to the response type of feedback (p=.000 in both cases).

Table 18 Response and the occurrence of uptake

	Number of episodes	Occurrence of uptake		
		Uptake	No uptake	No opportunity
Provide	1321	824 (62.4%)	304 (23.0%)	193 (14.6%)
Residuals (Provide)	--	-7.7	5.3	4.4
Elicit	113	111 (98.2%)	2 (1.8%)	0 (0.0%)

$X^2(2, n=1434)=59.045, p=.000$

Table 19 Response and the successfulness of uptake

	Number of uptake	Successfulness of uptake			
		Successful	Unsuccessful	Acknowledge	Inconclusive
Provide	824	319 (38.7%)	15 (1.8%)	349 (42.4%)	141 (17.1%)
Residuals (Provide)	--	-10.9	.0	8.1	4.2
Elicit	111	104 (93.7%)	2 (1.8%)	3 (2.7%)	2 (1.8%)

$X^2(3, n=935)=121.207, p=.000$

Linguistic focus

Tables 20, 21, 22, and 23 illustrate the occurrence and successfulness of learner uptake according to the linguistic focus of feedback episodes. Table 20 shows that learners responded to teacher feedback in 60.7% of vocabulary episodes, 68.8% of grammar episodes, 69.1% of pronunciation episode, and 84.2% of spelling episode. Adjusted standardized residuals in Table 21 reveal that uptake was significantly more frequent in grammar episodes while no opportunity was significantly more frequent in vocabulary episodes. Spelling episodes also resulted in significantly more uptake and meanwhile were least likely to result in no uptake. Table 22 shows that when uptake did occur, 38.4% was successful in vocabulary episodes, 42.6% was successful in grammar episodes, 85.7% was successful in pronunciation episodes, and 34.4% was successful in spelling episodes. Adjusted standardized residuals in Table 23 reveal that pronunciation episodes led to significantly more successful uptake, vocabulary led to significantly more

acknowledge, and spelling led to significantly more inconclusive uptake. There is no significant difference among the four in terms of unsuccessful uptake. Chi-square tests indicate that both the occurrence and successfulness of uptake are significantly related to the linguistic focus of feedback episodes (p=.000 in both cases).

Table 20 Linguistic focus and the occurrence of uptake

	Number of episodes	Occurrence of uptake		
		Uptake	No uptake	No opportunity
Vocabulary	712	432 (60.7%)	160 (22.5%)	120 (16.9%)
Grammar	532	366 (68.8%)	118 (22.2%)	48 (9.0%)
Pronunciation	152	105 (69.1%)	25 (16.4%)	22 (14.5%)
Spelling	38	32 (84.2%)	3 (7.9%)	3 (7.9%)

$X^2(6, n=1434)=26.145, p=.000$

Table 21 Linguistic focus and the occurrence of uptake residuals

	Uptake	No uptake	No opportunity
Vocabulary	-3.6	1.0	3.7
Grammar	2.2	.6	-3.8
Pronunciation	1.1	-1.6	.4
Spelling	2.5	-2.1	-1.0

Table 22 Linguistic focus and the successfulness of uptake

	Number of uptake	Successfulness of uptake			
		Successful	Unsuccessful	Acknowledge	Inconclusive
Vocabulary	432	166 (38.4%)	5 (1.1%)	196 (45.4%)	65 (15.0%)
Grammar	366	156 (42.6%)	10 (2.8%)	140 (38.3%)	60 (16.4%)
Pronunciation	105	90 (85.7%)	2 (1.9%)	13 (12.4%)	0 (0.0%)
Spelling	32	11 (34.4%)	0 (0.0%)	3 (9.4%)	18 (56.3%)

$X^2(9, n=935)=130.089, p=.000$

Table 23 Linguistic focus and the successfulness of uptake residuals

	Successful	Unsuccessful	Acknowledge	Inconclusive
Vocabulary	-3.9	-1.4	4.5	-.2
Grammar	-1.3	1.7	.3	.7
Pronunciation	8.8	.1	-5.7	-4.6
Spelling	-1.3	-.8	-3.4	6.5

Directness

Tables 24 and 25 illustrate the occurrence and successfulness of learner uptake according to the directness of feedback episodes. Table 24 shows that learners responded to teacher feedback in 62.0% of direct episodes and 76.1% of indirect episodes. Adjusted standardized residuals in the table reveal that uptake was significantly more frequent in indirect episodes while no uptake was significantly more frequent in direct episodes. There is no significant difference between the two in terms of no opportunity. Table 25 shows that when uptake did occur, 33.2% was successful in direct episodes and 78.3% was successful in indirect episodes. Adjusted standardized residuals in the table reveal that indirect episodes led to significantly more successful uptake while direct episodes led to significantly more acknowledge and inconclusive uptake. There is no significant difference between the two in terms of unsuccessful uptake. Chi-square tests indicate that both the occurrence and successfulness of uptake are significantly related to the directness of feedback episodes (p=.000 in both cases).

Table 24 Directness and the occurrence of uptake

	Number of episodes	Occurrence of uptake		
		Uptake	No uptake	No opportunity
Direct	1107	686 (62.0%)	269 (24.3%)	152 (13.7%)
Residuals (Direct)	--	-3.9	4.3	.4
Indirect	327	249 (76.1%)	37 (11.3%)	41 (12.5%)

$X^2(2, n=1434)=19.835, p=.000$

Table 25 Directness and the successfulness of uptake

	Number of uptake	Successfulness of uptake			
		Successful	Unsuccessful	Acknowledge	Inconclusive
Direct	686	228 (33.2%)	12 (1.7%)	309 (45.0%)	137 (20.0%)
Residuals (Direct)	--	-11.5	-1.4	7.5	6.5
Indirect	249	195 (78.3%)	5 (2.0%)	43 (17.3%)	6 (2.4%)

$X^2(3, n=935)=145.059, p=.000$

Complexity

Tables 26 and 27 illustrate the occurrence and successfulness of learner uptake according to the complexity of feedback episodes. Table 26 shows that learners responded to teacher feedback in 57.8% of simple episodes and 70.2% of complex episodes. Adjusted standardized residuals in the table reveal that uptake was significantly more frequent in complex episodes while no opportunity was significantly more frequent in simple episodes. There is no significant difference between the two in terms of no uptake. Table 27 shows that when uptake did occur, 38.1% was successful in simple episodes and 49.2% was successful in complex episodes. Adjusted standardized residuals in the table reveal that complex episodes led to both significantly more successful and unsuccessful uptake while simple episodes led to significantly more acknowledge. There is no significant difference between the two in terms of inconclusive uptake. Chi-square tests indicate that both the occurrence and successfulness of uptake are significantly related to the complexity of feedback episodes (p value is .000 for the occurrence of uptake and .001 for the successfulness of uptake).

Table 26 Complexity and the occurrence of uptake

	Number of episodes	Occurrence of uptake		
		Uptake	No uptake	No opportunity
Simple	573	331 (57.8%)	124 (21.6%)	118 (20.6%)
Residuals (Simple)	--	-4.8	.2	6.5
Complex	861	604 (70.2%)	182 (21.1%)	75 (8.7%)

$X^2(2, n=1434)=44.227, p=.000$

Table 27 Complexity and the successfulness of uptake

	Number of uptake	Successfulness of uptake			
		Successful	Unsuccessful	Acknowledge	Inconclusive
Simple	331	126 (38.1%)	2 (0.6%)	149 (45.0%)	54 (16.3%)
Residuals (Simple)	--	-3.3	-2.1	3.4	.6
Complex	604	297 (49.2%)	15 (2.5%)	203 (33.6%)	89 (14.7%)

$X^2(3, n=935)=17.720, p=.001$

Emphasis

Tables 28 and 29 illustrate the occurrence and successfulness of learner uptake according to the emphasis of feedback episodes. Table 28 shows that learners responded to teacher feedback in 61.0% of light episodes and 65.8% of heavy episodes. Adjusted standardized residuals in the table reveal that there is no significant difference between the two in terms of uptake and no uptake but no opportunity was significantly more frequent in light episodes. Table 29 shows that when uptake did occur, 64.0% was successful in light episodes and 42.7% was successful in heavy episodes. Adjusted standardized residuals in the table reveal that light episodes led to significantly more successful uptake while heavy episodes led to significantly more inconclusive uptake. There is no significant difference between the two in terms of unsuccessful uptake and acknowledge. Chi-square tests indicate that both the occurrence and successfulness of uptake are significantly

related to the emphasis of feedback episodes (p value is .014 for the occurrence of uptake and .000 for the successfulness of uptake)

Table 28 Emphasis and the occurrence of uptake

	Number of episodes	Occurrence of uptake		
		Uptake	No uptake	No opportunity
Light	182	111 (61.0%)	34 (18.7%)	37 (20.3%)
Residuals (Light)	--	-1.3	-.9	2.9
Heavy	1252	824 (65.8%)	272 (21.7%)	156 (12.5%)

$X^2(2, n=1434)=8.569, p=.014$

Table 29 Emphasis and the successfulness of uptake

	Number of uptake	Successfulness of uptake			
		Successful	Unsuccessful	Acknowledge	Inconclusive
Light	111	71 (64.0%)	2 (1.8%)	35 (31.5%)	3 (2.7%)
Residuals (Light)	--	4.2	.0	-1.4	-3.9
Heavy	824	352 (42.7%)	15 (1.8%)	317 (38.5%)	140 (17.0%)

$X^2(3, n=935)=24.068, p=.000$

To sum up, the descriptive and inferential analysis in this section reveals that all the seven characteristics of feedback episodes significantly affected both the occurrence and successfulness of uptake ($p<.05$). Adjusted standardized residuals indicate that code-related, elicit, indirect, and complex episodes resulted in both significantly more uptake and significantly more successful uptake (ASR>2.0). In addition to episodes with these characteristics, light episodes also resulted in significantly more successful uptake. Regarding the type and linguistic focus of feedback episodes, while student-initiated and spelling episodes led to significantly more learner uptake, reactive and pronunciation episodes led to significantly more successful uptake.

4.4 Characteristics of feedback episodes and recall comments

This section reports the relationship between the general characteristics of feedback episodes and learners' comments in the stimulated recall interviews. After an overview of the overall recall comments, the frequency of comments related to teacher feedback is presented in relation to each characteristic in the order of type, source, response, linguistic focus, directness, complexity, and emphasis.

Overview of recall comments

Tables 30, 31, and 32 show the nature of learners' comments during the stimulated recall interviews and therein their attention to teacher feedback. As Table 30 shows, among a total of 279 episodes, learners were thinking about language issues 86.0% of the time. In 6.5% of the episodes they were thinking about non-language-related issues. In 3.9% of the episodes they were not thinking of anything. For the rest of the episodes, they either didn't remember what they were thinking or made a comment that was impossible to classify. Table 31 shows that among a total of 240 language-related comments, 97.5% are target-feature-related, 1.6% are non-target-feature-related, and 0.8% are inconclusive. Table 32 specifically shows learners' attention to teacher feedback. Among a total of 234 target-feature-related comments, 85.9% are related to teacher feedback, 12.8% are not related to teacher feedback, and the rest are inconclusive. While the different tiers of analysis provide an overview of all the recall comments, only the last one directly addresses learners' noticing of teacher feedback, one of the two major objectives of the study. The statistical analysis of recall comments therefore only focused on target-feature-related comments. Meanwhile, other comments, namely, non-language-related

comments, no thought, no memory, unclassifiable comments, non-target-feature-related comments, and inconclusive language-related comments, compose an important part of the recall interviews. Instead of totally excluding these categories from the chi-square analysis, they were all considered together with inconclusive target-feature-related comments as one category: other.

Table 30 Distribution of overall comments

Total number of comments	Language-related	Non-language-related	No thought	No memory	Unclassifiable
279	240 (86.0%)	18 (6.5%)	11 (3.9%)	7 (2.5%)	3 (1.1%)

Table 31 Distribution of language-related comments

Total number of language-related comments	Target-feature-related	Non-target-feature-related	Inconclusive
240	234 (97.5%)	4 (1.6%)	2 (0.8%)

Table 32 Distribution of target-feature-related comments

Total number of target-feature-related comments	Teacher-feedback-related	Non-teacher-feedback-related	Inconclusive
234	201 (85.9%)	30 (12.8%)	3 (1.3%)

Type

Tables 33 and 34 show the rate of teacher-feedback-related comments in relation to the type of feedback episodes. Table 33 shows that 74.8% of recall comments are related to teacher feedback for teacher-initiated episodes, 76.0% are related to teacher feedback for student-initiated episodes, and 60.9% are related to teacher feedback for reactive episodes. Adjusted standardized residuals in Table 34 indicate that learners talked more (although not significantly) about teacher feedback after viewing student-initiated episodes and

significantly more about other issues after viewing reactive episodes. Chi-square test results suggest that the rate of teacher-feedback-related comments was significantly affected by the type of feedback episodes (p=.041).

Table 33 Type and teacher-feedback-related comments

	Total number of comments	Teacher-feedback-related	Non-teacher-feedback-related	Other
T-initiated	115	86 (74.8%)	12 (10.4%)	17 (14.8%)
S-initiated	100	76 (76.0%)	12 (12.0%)	12 (12.0%)
Reactive	64	39 (60.9%)	6 (9.4%)	19 (29.7%)

$X^2(4, n=279)=9.988, p=.041$

Table 34 Type and teacher-feedback-related comments residuals

	Teacher-feedback-related	Non-teacher-feedback-related	Other
T-initiated	.9	-.1	-.9
S-initiated	1.2	.5	-1.8
Reactive	-2.4	-.4	3.1

Source

Table 35 shows the rate of teacher-feedback-related comments in relation to the source of feedback episodes. A total of 68.1% of recall comments are related to teacher feedback for code-related episodes and 76.7% are related to teacher feedback for message-related episodes. Chi-square test results suggest that the rate of teacher-feedback-related comments was not significantly affected by the source of feedback episodes (p=.198).

Table 35 Source and teacher-feedback-related comments

	Total number of comments	Teacher-feedback-related	Non-teacher-feedback-related	Other
Code	144	98 (68.1%)	17 (11.8%)	29 (20.1%)
Message	135	103 (76.7%)	13 (10.0%)	19 (13.3%)

$X^2(2, n=279)=3.240$, $p=.198$ (Semantic episodes were conflated with message-related episodes due to low expected cell counts)

Response

Table 36 shows the rate of teacher-feedback-related comments in relation to the response type of feedback episodes. A total of 72.0% of recall comments are related to teacher feedback for provide episodes and 72.2% are related to teacher feedback for elicit episodes. The two are virtually the same. Chi-square test results suggest that the rate of teacher-feedback-related comments was not significantly affected by the response type of feedback episodes ($p=.997$).

Table 36 Response and teacher-feedback-related comments

	Total number of comments	Teacher-feedback-related	Non-teacher-feedback-related	Other
Provide	261	188 (72.0%)	28 (10.7%)	45 (17.2%)
Elicit	18	13 (72.2%)	2 (11.1%)	3 (16.7%)

$X^2(2, n=279)=.006$, $p=.997$

Linguistic focus

Table 37 shows the rate of teacher-feedback-related comments in relation to the linguistic focus of feedback episodes. A total of 76.0% of recall comments are related to teacher feedback for vocabulary episodes, 63.1% are related to teacher feedback for grammar episodes, and 70.0% are related to teacher feedback for pronunciation episodes. Chi-

square test results suggest that the rate of teacher-feedback-related comments was not significantly affected by the linguistic focus of feedback episodes (p=.227).

Table 37 Linguistic focus and teacher-feedback-related comments

	Total number of comments	Teacher-feedback-related	Non-teacher-feedback-related	Other
Vocabulary	175	133 (76.0%)	18 (10.3%)	24 (13.7%)
Grammar	65	41 (63.1%)	7 (10.8%)	17 (26.2%)
Pronunciation	30	21 (70.0%)	4 (13.3%)	5 (16.7%)

$X^2(4, n=270)=5.648$, p=.227 (Spelling was excluded due to low expected cell counts)

Directness

Table 38 shows the rate of teacher-feedback-related comments in relation to the directness of feedback episodes. A total of 73.5% of recall comments are related to teacher feedback for direct episodes and 61.8% are related to teacher feedback for indirect episodes. Adjusted standardized residuals indicate that learners talked more (although not significantly) about teacher feedback after viewing direct episodes and significantly more about other issues after viewing indirect episodes. Chi-square test results suggest that the rate of teacher-feedback-related comments was significantly affected by the directness of feedback episodes (p=.029).

Table 38 Directness and teacher-feedback-related comments

	Total number of comments	Teacher-feedback-related	Non-teacher-feedback-related	Other
Direct	245	180 (73.5%)	28 (11.4%)	37 (15.1%)
Residuals (Direct)	--	1.6	.9	-2.6
Indirect	34	21 (61.8%)	2 (5.9%)	11 (32.4%)

$X^2(2, n=279)=7.105$, p=.029

Complexity

Table 39 shows the rate of teacher-feedback-related comments in relation to the complexity of feedback episodes. A total of 62.1% of recall comments are related to teacher feedback for simple episodes and 74.4% are related to teacher feedback for complex episodes. Adjusted standardized residuals indicate that learners talked significantly more about teacher feedback after viewing complex episodes and significantly more about other issues after viewing simple episodes. Chi-square test results suggest that the rate of teacher-feedback-related comments was significantly affected by the complexity of feedback episodes (p=.049).

Table 39 Complexity and teacher-feedback-related comments

	Total number of comments	Teacher-feedback-related	Non-teacher-feedback-related	Other
Simple	58	36 (62.1%)	6 (10.3%)	16 (27.6%)
Residuals (Simple)	--	-2.0	-.1	2.4
Complex	221	165 (74.7%)	24 (10.9%)	32 (14.5%)

$X^2(2, n=279)=6.045, p=.049$

Emphasis

Table 40 shows the rate of teacher-feedback-related comments in relation to the emphasis of feedback episodes. A total of 53.8% of recall comments are related to teacher feedback for light episodes and 72.9% are related to teacher feedback for heavy episodes. Adjusted standardized residuals indicate that learners talked more (although not significantly) about teacher feedback after viewing heavy episodes and significantly more about other issues after viewing light episodes. Chi-square test results suggest that the rate of teacher-

feedback-related comments was significantly affected by the emphasis of feedback episodes (p=.007).

Table 40 Emphasis and teacher-feedback-related comments

	Total number of comments	Teacher-feedback-related	Non-teacher-feedback-related	Other
Light	13	7 (53.8%)	0 (0.0%)	6 (46.2%)
Residuals (Light)	--	-1.7	-1.2	3.1
Heavy	266	194 (72.9%)	30 (11.3%)	42 (15.8%)

$X^2(2, n=279)=10.033, p=.007$

To sum up, the descriptive and inferential analysis in this section reveals that among the seven characteristics of feedback episodes, only the type, directness, complexity, and emphasis of feedback episodes significantly affected the rate of teacher-feedback-related comments ($p<.05$). In other words, source, response, and linguistic focus did not significantly affect the rate of teacher-feedback-related comments. Adjusted standardized residuals indicate that complex episodes led to significantly more teacher-feedback-related comments. Learners also made more comments about teacher feedback after viewing student-initiated, direct, and heavy episodes although the values of the residuals are not up to 2.0.

4.5 Characteristics of feedback episodes and test results

This section reports the relationship between the general characteristics of feedback episodes and test results. After an overview of the general test results, results from both the immediate test and delayed test are presented together according to each characteristic in the order of type, source, response, linguistic focus, directness, complexity, and emphasis.

Overview of test results

Table 41 shows the general results of the two tests. In the immediate test, learners provided a correct answer for 64.7% of the total items, an incorrect answer for 30.9% of the total items, and a partially correct answer for 4.4% of the total items. In the delayed test, the percentage of correct answers (53.2%) declined by 11.5% while the percentages of incorrect answers (42.2%) and partially correct answers (4.6%) increased by 11.3% and 0.2% respectively.

Table 41 Overview of test results

	Immediate test		Delayed test	
	Number	Percentage	Number	Percentage
Correct	176	64.7%	150	53.2%
Incorrect	84	30.9%	119	42.2%
Partially correct	12	4.4%	13	4.6%
Total	272	100%	282	100%

Type

Table 42 is an illustration of test results according to the type of feedback episodes. In the immediate test, the percentage of correct answers is 69.0% for teacher-initiated episodes, 72.9% for student-initiated episodes, and 56.0% for reactive episodes. Adjusted standardized residuals suggest that learners provided significantly more correct answers for test items developed from student-initiated episodes and significantly less correct answers for items developed from reactive episodes. In the delayed test, the percentages of correct answers for the three types of feedback episodes are very close to one another, with 55.7% for teacher-initiated episodes, 50.6% for student-initiated episodes, and

52.3% for reactive episodes. Chi-square analysis indicates that the type of feedback episodes significantly affected the results of the immediate test (p=.030) but not the results of the delayed test (p=.772).

Table 42 Type and test results

	Immediate test				Delayed test		
	Number of test items	Correct	Residuals (Correct)	Incorrect	Number of test items	Correct	Incorrect
T-initiated	71	49 (69.0%)	.8	22 (31%)	115	64 (55.7%)	51 (44.4%)
S-initiated	85	62 (72.9%)	2.0	23 (27.1%)	79	40 (50.6%)	39 (49.4%)
Reactive	116	65 (56.0%)	-2.6	51 (43.9%)	88	46 (52.3%)	42 (47.7%)
	$X^2(2, n=272)=7.030, p=.030$				$X^2(2, n=282)=.517, p=.772$		

Source

Table 43 is an illustration of test results according to the source of feedback episodes. In the immediate test, the percentage of correct answers is 63.3% for code-related episodes, 67.4% for message-related episodes, and 66.7% for semantic episodes. In the delayed test, the percentage of correct answers is 55.4% for code-related episodes and 49.5% for message-related episodes. Chi-square analysis indicates that the source of feedback episodes did not significantly affect the results of either test (p value is .796 for the immediate test and .342 for the delayed test).

Table 43 Source and test results

	Immediate test			Delayed test		
	Number of test items	Correct	Incorrect	Number of test items	Correct	Incorrect
Code	177	112 (63.3%)	65 (36.8%)	177	98 (55.4%)	79 (44.6%)
Message	86	58 (67.4%)	28 (32.5%)	105	52 (49.5%)	53 (50.5%)
Semantic	9	6 (66.7%)	3 (33.3%)	Conflated with message due to low expected cell counts		
	$X^2(2, n=272)=.455, p=.796$			$X^2(1, n=282)=.904, p=.342$		

Response

Table 44 is an illustration of test results according to the response type of feedback episodes. In the immediate test, the percentage of correct answers is 63.7% for provide episodes and 75.0% for elicit episodes. In the delayed test, the percentage of correct answers for the two is 51.9% and 70.0% respectively. Chi-square analysis indicates that the response type of feedback episodes did not significantly affect the results of either test (*p* value is .269 for the immediate test and .118 for the delayed test).

Table 44 Response and test results

	Immediate test			Delayed test		
	Number of test items	Correct	Incorrect	Number of test items	Correct	Incorrect
Provide	248	158 (63.7%)	90 (36.3%)	262	136 (51.9%)	126 (48.1%)
Elicit	24	18 (75.0%)	6 (25.0%)	20	14 (70.0%)	6 (30.0%)
	$X^2(1, n=272)=1.221, p=.269$			$X^2(1, n=282)=2.443, p=.118$		

Linguistic focus

Table 45 is an illustration of test results according to the linguistic focus of feedback episodes. In the immediate test, the percentage of correct answers is 68.8% for vocabulary episodes, 67.0% for grammar episodes, and 48.7% for pronunciation episodes.

In the delayed test, the percentage of correct answers for the three is 46.8%, 64.2%, and 52.9% respectively. The percentage of correct answers for spelling episodes is 26.7%. Adjusted standardized residuals suggest that learners provided significantly more correct answers for test items developed from grammar episodes and significantly less correct answers for items developed from spelling episodes. Chi-square analysis indicates that the linguistic focus of feedback episodes significantly affected the results of the delayed test (p=.013) but not the results of the immediate test (p=.063).

Table 45 Linguistic focus and test results

	Immediate test			Delayed test			
	Number of test items	Correct	Incorrect	Number of test items	Correct	Residuals (Correct)	Incorrect
Vocabulary	125	86 (68.8%)	39 (31.2%)	124	58 (46.8%)	-1.8	66 (53.3%)
Grammar	103	69 (67.0%)	34 (33.1%)	109	70 (64.2%)	2.8	39 (35.8%)
Pronunciation	39	19 (48.7%)	20 (51.3%)	34	18 (52.9%)	.0	16 (47%)
Spelling	Excluded due to low expected cell counts			15	4 (26.7%)	-2.1	11 (73.3%)
	$X^2(2, n=267)=5.526, p=.063$			$X^2(3, n=282)=10.860, p=.013$			

Directness

Table 46 is an illustration of test results according to the directness of feedback episodes. In the immediate test, the percentage of correct answers is 68.4% for direct episodes and 53.0% for indirect episodes. Adjusted standardized residuals suggest that learners provided significantly more correct answers for test items developed from direct episodes. In the delayed test, the percentage of correct answers is 54.3% for direct episodes and 49.2% for indirect episodes. Chi-square analysis indicates that the directness of feedback

episodes significantly affected the results of the immediate test (p=.023) but not the results of the delayed test (p=.484).

Table 46 Directness and test results

	Immediate test				Delayed test		
	Number of test items	Correct	Residuals (Correct)	Incorrect	Number of test items	Correct	Incorrect
Direct	206	141 (68.4%)	2.3	65 (31.5%)	223	121 (54.3%)	102 (45.7%)
Indirect	66	35 (53.0%)	-2.3	31 (47.0%)	59	29 (49.2%)	30 (50.9%)
	$X^2(1, n=272)=5.202, p=.023$				$X^2(1, n=282)=.489, p=.484$		

Complexity

Table 47 is an illustration of test results according to the complexity of feedback episodes. In the immediate test, the percentage of correct answers is 54.5% for simple episodes and 69.6% for complex episodes. In the delayed test, the percentage of correct answers for the two is 45.5% and 58.2% respectively. Adjusted standardized residuals suggest that learners provided significantly more correct answers for test items developed from complex episodes in both the immediate test and delayed test. Chi-square analysis indicates that the complexity of feedback episodes significantly affected the results of both tests (p value is .015 and .037 respectively).

Table 47 Complexity and test results

	Immediate test				Delayed test			
	Number of test items	Correct	Residuals (Correct)	Incorrect	Number of test items	Correct	Residuals (Correct)	Incorrect
Simple	88	48 (54.5%)	-2.4	40 (45.5%)	112	51 (45.5%)	-2.1	61 (54.5%)
Complex	184	128 (69.6%)	2.4	56 (30.4%)	170	99 (58.2%)	2.1	71 (41.8%)
	$X^2(1, n=272)=5.880, p=.015$				$X^2(1, n=282)=4.374, p=.037$			

Emphasis

Table 48 is an illustration of test results according to the emphasis of feedback episodes. In the immediate test, the percentage of correct answers is 32.3% for light episodes and 68.9% for heavy episodes. Adjusted standardized residuals suggest that learners provided significantly more correct answers for test items developed from heavy episodes. In the delayed test, the percentage of correct answers is 41.2% for light episodes and 54.8% for heavy episodes. Chi-square analysis indicates that the emphasis of feedback episodes significantly affected the results of the immediate test (p=.000) but not the results of the delayed test (p=.134).

Table 48 Emphasis and test results

	Immediate test				Delayed test		
	Number of test items	Correct	Residuals (Correct)	Incorrect	Number of test items	Correct	Incorrect
Light	31	10 (32.3%)	-4.0	21 (67.7%)	34	14 (41.2%)	20 (58.8%)
Heavy	241	166 (68.9%)	4.0	75 (31.2%)	248	136 (54.8%)	112 (45.1%)
	$X^2(1, n=272)=16.130, p=.000$				$X^2(1, n=282)=2.242, p=.134$		

To sum up, the descriptive and inferential analysis in this section reveals that the results of the immediate test were significantly affected by the type, directness, complexity, and emphasis of feedback episodes ($p<.05$). Adjusted standardized residuals indicate that learners provided significantly more correct answers for items developed from student-initiated, direct, complex, and heavy episodes (ASR>2.0) than those from teacher-initiated/reactive, indirect, simple, and light episodes. The results of the delayed test, on the other hand, were significantly affected only by the linguistic focus and complexity of feedback episodes. Adjusted standardized residuals indicate that learners provided

significantly more correct answers for items developed from grammar and complex episodes than those from vocabulary/pronunciation/spelling and simple episodes.

4.6 Review of results

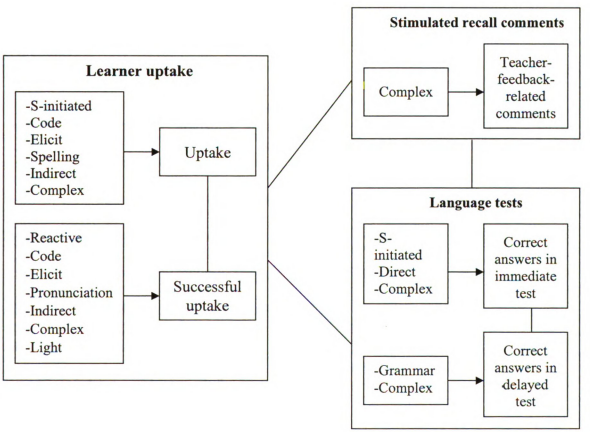
The results presented above show that all seven characteristics of feedback episodes had a statistically significant bearing on the occurrence and successfulness of learner uptake. In terms of teacher-feedback-related recall comments, however, only type, directness, complexity, and emphasis showed a statistically significant effect. These four characteristics were also shown to have significantly affected the results of the immediate test. Results of the delayed test, on the other hand, were only significantly influenced by linguistic focus and complexity. The table below is a summary of the overall results in this section. The plus sign symbolizes a statistically significant relationship while the minus sign represents a statistically non-significant relationship. This convention also applies to the summary tables in Chapters 5 and 6.

Table 49 Overall results by the characteristics of feedback episodes

	Occurrence of uptake	Successfulness of uptake	Feedback-related comments	Immediate test	Delayed test
Type	+	+	+	+	-
Source	+	+	-	-	-
Response	+	+	-	-	-
Linguistic focus	+	+	-	-	+
Directness	+	+	+	+	-
Complexity	+	+	+	+	+
Emphasis	+	+	+	+	-

Figure 5 is an overview of the characteristics that predicted uptake, successful uptake, teacher-feedback-related comments, and correct test results. Specifically, student-initiated, code-related, elicit, spelling, indirect, and complex episodes led to more uptake. Reactive, code-related, elicit, pronunciation, indirect, complex, and light episodes led to more successful uptake. Teacher-feedback-related comments were only predicted by complex episodes. As for test results, more correct answers resulted from student-initiated, direct, complex, and heavy episodes than from teacher-initiated/reactive, indirect, simple, and light episodes in the immediate test and more from grammar and complex episodes than from vocabulary/pronunciation/spelling and simple episodes in the delayed test.

Figure 5 Overview of the characteristics of feedback episodes predicting (successful) uptake, teacher-feedback-related comments, and correct test results



4.7 Discussion

This section discusses the results presented in the previous sections. The occurrence and successfulness of uptake are examined together with learners' stimulated recall comments and test results to explore learners' noticing of teacher feedback and the consequent language improvement. Due to the cell count requirement of chi-square tests, not all the categories of each variable were included in the results. Some categories are therefore discussed solely according to descriptive statistics. On the other hand, an examination of every category regardless of their relative importance would make the discussion extremely long and tedious. The discussion therefore focuses on categories that stand out from the results. Also, in the effort to interpret the results, new calculations were made when necessary. These rules also apply to the discussion in Chapters 5 and 6 (in Chapter 5, there are also variables which did not receive chi-square analysis due to low expected cell counts).

One thing the present study is different from previous studies examining the characteristics of feedback episodes (e.g., Ellis et al., 2001a, 2001b; Loewen, 2004, 2005) is the operationalization of uptake and the classification of the successfulness of uptake. In those studies, uptake only included learners' verbal responses to teacher feedback. It was classified as either successful or unsuccessful in inferential statistical analysis, with unsuccessful uptake covering other types of uptake such as acknowledgement or inconclusive uptake. In the present study, the use of a video camera made it possible for me to look at learners' nonverbal responses to teacher feedback, for example, nodding, writing, acting, and pointing (See chapter 3 for detail). Acting and pointing occurred in a

small number of cases and were coded as successful uptake as I saw appropriate. Nodding and writing alone (i.e., not including those accompanied by verbal responses such as "oh" and "yeah"), on the other hand, occurred frequently across the nine classes. Nodding was coded as acknowledge and writing was coded as inconclusive uptake. Consequently, there is a high percentage of both acknowledge (37.6%) and inconclusive uptake (15.3%). Given the big percentage of these response types and the possibility that they might make a difference to the overall results, it was decided that they should not be excluded from the statistical analysis and that they should receive sufficient attention in the discussion of the results.

Before scrutinizing the detailed results, it is necessary to take a look at the overall results. Among a total of 1434 episodes, 65.2% resulted in some kind of uptake and 45.2% of the uptake was successful. In Ellis et. al (2001a) and Loewen's (2004) studies, the level of both uptake and successful uptake was above or close to 70%. One possible explanation for the lower level of uptake in the present study is the opportunities learners were given to respond to teachers. In the study of Ellis et al. (2001a), no opportunity was not reported in the result. It is not clear whether it was coded as no uptake or not included at all. In Loewen's study, the overall percentage of no opportunity was 9.5%, 4% lower than that of the present study (13.5%). This means learners in the present study had fewer opportunities to respond to teachers. It is possible that they would produce a higher level of uptake if given more opportunities. As for the lower level of successful uptake, it could be because the inclusion of nonverbal uptake in the present study largely increased

the amount of acknowledge and inconclusive uptake, which in turn lowered the percentage of successful uptake.

Regarding the effect of specific characteristics on the occurrence and successfulness of uptake, the three studies are not identical in the number of characteristics examined. In addition to all the seven characteristics in the present study, Loewen also examined the timing of uptake (i.e., when teacher feedback is given). Ellis et al., on the other hand, only looked at type, source, directness, complexity, and linguistic focus. Moreover, the design of the studies is not exactly the same either. In addition to the differences described in the previous paragraph, the object of examination is different too. Both Ellis et al. and Loewen's studies explored the effect of focus on form. Activities with a focus on pre-targeted forms were excluded from the analysis. The episodes they examined therefore were strictly meaning-based. The present study, in contrast, investigated the effect of feedback episodes, which as illustrated in Chapter 1 is more of an umbrella concept. Activities which aimed to promote language accuracy were included in the analysis. The episodes examined therefore are partially meaning-focused and partially form-focused. Due to these differences, it is not easy to compare the specifics about each characteristic. One thing these studies do have in common is that some characteristics, complexity, for instance, have a significant effect on the occurrence and successfulness of uptake.

With respect to testing, no test was reported in either Ellis et al. or Loewen's study. Another study by Loewen (2005), however, did report the test results based on the same

data as his 2004 study. The overall percentage of correct answers was 43.3%, which was 15.5% lower than the percentage in the present study (58.8%). The higher percentage of correct answers in the present study could be a result of the inclusion of test items developed from form-focused activities. Meanwhile, both studies found that the percentage of correct answers in general was lower in the delayed test than in the immediate test (dropped by 8.3% in Lowen's study and 11.5% in the present study). This is a natural result of memory decay. Finally, concerning the stimulated recall comments, none of the other three studies used it as an instrument. It is carefully examined in the detailed discussion of each characteristic below.

The first characteristic examined is type. It describes how a feedback episode is initiated. Chi-square analysis indicates that the type of feedback episodes significantly affected both the occurrence and successfulness of uptake (p=.000 in both cases). In terms of uptake, student-initiated episodes resulted in the highest level of uptake (71.6%) while teacher-initiated episodes led to the lowest level of uptake (54.7%), with reactive episodes in the middle (69.6%). This result could have been caused by the nature of the language problems in the episodes. In student-initiated episodes, the questions learners asked were real problems learners thought they needed to know. Learners therefore might be more responsive to teachers' feedback. In teacher-initiated episodes, however, the problems were based on teachers' assessment of learners' knowledge. Learners might not have these problems at all. They therefore might be less active in their interactions with teachers. This could be the reason why teacher-initiated episodes also have a significantly higher level of no uptake (ASR=2.6) than the other two types of episodes. In reactive

episodes, even when there was a problem, learners might not feel the urgency to attend to it. They therefore might not take as much initiative to respond to the feedback either. Given these differences of the problems involved in the different types of episodes, it is not strange that student-initiated episodes resulted in more uptake.

The initiation learners might take in student-initiated episodes could lead one to expect more successful uptake in these episodes too. This nevertheless is not true. It turned out that reactive episodes resulted in the highest level of successful uptake while student-initiated episodes led to the lowest level of successful uptake (ASR value is 6.5 and -6.6 respectively). Student-initiated episodes also resulted in the highest level of acknowledge (ASR=6.5). A look at the specifics of student-initiated episodes and reactive episodes reveals part of the possible reasons why this is so. In a lot of student-initiated episodes (51.6%), learners asked about the meaning of a word or asked about a word they needed to express themselves. More often than not, they simply acknowledged the feedback once they could continue with their interactions with the teacher or peers or when they could continue the task at hand, hence leading to more acknowledge and less successful uptake. Different from student-initiated episodes, a good percentage of reactive episodes (58.6%) involve pronunciation problems. In these episodes, teachers often provided the correct pronunciation and learners often simply repeated teachers right after the feedback, leading to more successful uptake (see more in the discussion of linguistic focus).

The type of feedback episodes was also found to be significantly related to teacher-feedback-related comments (p=.041). To be specific, student-initiated episodes led to the

highest percentage of teacher-feedback-related comments (76.0%) while reactive episodes led to the lowest percentage of teacher-feedback-related comments (60.9%). This again could be explained by the reasons given above -- the nature of the problems involved in the feedback episodes and how learners viewed it. That is, learners might pay more attention to the feedback about a problem they were experiencing in student-initiated episodes and less attention to the feedback about a problem initiated or detected by other people, in this case, the teacher. As for the reason why reactive episodes, not teacher-initiated episodes, led to the lowest percentage of teacher-feedback-related comments, it could be because learners thought a question initiated by the teacher might be of some importance and therefore paid more attention to it whether they understood it or not.

Concerning test results, chi-square analysis indicates that there is a statistically significant difference among the three types of feedback episodes in the immediate test (p=.031). Learners provided significantly more correct answers for test items developed from student-initiated episodes (ASR=2.0) and significantly fewer correct answers for items developed from reactive episodes (ASR=-2.4). This result corresponds well with the results from the stimulated recall interviews -- more teacher-feedback-related comments associated with student-initiated episodes and fewer teacher-feedback-related comments associated with reactive episodes. The heavier attention to teacher feedback in student-initiated episodes could have resulted in better performance with items developed from these episodes. As for the delayed test, it was not significantly affected by the type of

feedback episodes, probably because student-initiated episodes lost their advantage as a result of memory decay.

The second characteristic examined is source. It describes the origin of a feedback episode. Chi-square analysis indicates that the source of feedback episodes had a significant effect on both the occurrence and successfulness of uptake. Code-related episodes led to significantly more uptake (ASR=4.5) and more successful uptake (ASR=5.3) than message-related episodes (ASR value is -5.4 for uptake and -5.3 for successful uptake). Semantic episodes also led to significantly more uptake than message-related episodes, but only half of that of code-related episodes (ASR=2.2). Moreover, just like message-related episodes, it led to less successful uptake (ASR=-1.5) than code-related episodes as well. Semantic episodes are those where learners understood the overall discourse but did not know the meaning of particular words or phrases (Loewen, 2002). These episodes therefore are also meaning-based. If semantic episodes were to be treated as part of message-related episodes as Ellis et al. (2001a) did, it could be said that code-related episodes in general led to more (successful) learner responses than message-related episodes. This could be due to two reasons. First, in message-related episodes, learners were primarily concerned with communication. Once the meaning of a language form was clearly conveyed and the communication breakdown was fixed, learners would continue with their interactions with teachers or peers. This could be the reason why there are significantly more acknowledge (ASR=6.3) and no uptake (ASR=2.6) in message-related episodes. In code-related episodes, however, learners were primarily concerned with the accuracy of language forms. There was no

pressing need to continue a conversation and they could afford to uptake teacher feedback and try to get it right. Another possible explanation is that because this is an intensive English program and there were some grammar components in the majority of the classes observed, learners might want to make sure they understand the target features and would respond to teachers more (correctly) in episodes involving language forms.

According to what has been said, one might wonder whether learners would pay more attention to code-related feedback than message-related feedback. Analysis of the stimulated recall comments, however, shows that this is not the case. Although source was shown not to significantly affect the rate of teacher-feedback-related comments (p=.198), the percentage of teacher-feedback-related comments is 8.6% higher for message-related episodes than code-related episodes (76.7% and 68.1% respectively). It is hard to tell exactly why this is so, but it does warn us that uptake is an optional move (Ellis et al., 2001a; Loewen, 2004; Sheen, 2006), that less uptake does not necessarily mean less noticing, and that learners might attend to teacher feedback even if they simply acknowledge it or do not respond to it at all.

Regarding the test results, no statistically significant relationship was observed between source and test results (p>.05 in both the immediate test and delayed test). Given that there is no statistically significant relationship between source and teacher-feedback-related comments either, a tentative conclusion can be drawn that because source did not affect learners' noticing of teacher feedback, it did not affect their learning. A careful look at the percentages of correct answers in the immediate test shows that learners did

slightly better with test items developed from message-related episodes and semantic episodes than those developed from code-related episodes (67.4% and 66.7% versus 63.3%). In the delayed test, however, learners provided slightly more correct answers for items developed from code-related episodes than those from message-related episodes (55.4% versus 49.5%). One possible explanation for this change is that learners had been re-exposed to some language forms that had been treated in code-related episodes before the delayed test was administered. This is possible now that there were some focus grammar activities or review activities in the lessons observed. As for target features treated in message-related episodes, even though learners might also run across the same language forms again, the chance is slimmer now that the language foci of these episodes are generally more extensive and are not the focus of a lesson like grammar rules.

The third characteristic examined is response. It describes how teacher feedback is given. Chi-square analysis indicates that there is a statistically significant relationship between teachers' response type and the occurrence of uptake. Adjusted standardized residuals reveal that the level of uptake is significantly higher in elicit episodes (ASR=7.7) while the level of no uptake and no opportunity is significantly higher in provide episodes (ASR value is 5.3 and 4.4 respectively). There are two possible reasons why elicit episodes led to more learner responses. One is that when teachers elicited answers from learners, learners were "forced" to give an answer. The "forced" answer often constituted learner uptake. Even though learners were not able to provide an answer all the time, the expectation for them to do so largely increased the chance of learner uptake. Another reason might be that many of these learners, aged around 20, were anxious to impress the

teacher in elicit episodes. As one learner reported in a stimulated recall interview: "I want to let her know what I know...", learners often tried to get the right answer when they were given the opportunity to do so. This motivation to be the "knowing student" could have also contributed to the higher level of uptake in elicit episodes. When teachers provided answers, on the other hand, learners did not have to say anything if they did not want to. Neither were there so many opportunities for them to impress teachers. This might explain why there is a much higher percentage of no uptake (23.0%) and no opportunity (14.6%) in provide episodes.

In terms of the successfulness of uptake, chi-square analysis indicates that it was also significantly affected by teachers' response type. Adjusted standardized residuals reveal that elicit episodes resulted in significantly more successful uptake (ASR=10.9) while provide episodes resulted in significantly more acknowledge and inconclusive uptake (ASR value is 8.1 and 4.2 respectively). This result could also be explained by learners' positions in the two types of episodes. In elicit episodes, often times learners would manage to get the correct answers with the teachers' and sometimes classmates' help. The correct answers would be considered as successful uptake. In provide episodes, on the other hand, learners were often in a more passive position and might only give a simple response without striving to use the right language form. The fact that the simple provision of the correct form or language information as compared to elicitation led to a surprisingly higher percentage of acknowledge (42.4% as against 2.7%) and inconclusive uptake (17.1% as against 1.8%) could be evidence for this.

Given that elicit episodes led to both significantly more uptake and successful uptake, one might think they would also lead to significantly more teacher-feedback-related comments. Nevertheless, chi-square analysis indicates that there is no statistically significant difference between provide episodes and elicit episodes in terms of the recall comments. A possible explanation for this result is that when teachers tried to elicit answers from learners, learners took it as part of natural classroom interactions, not an effort to draw their attention to a particular language form; they therefore did not pay special attention to the feedback in elicit episodes. On the other hand, descriptive statistics reveal that elicit episodes did result in a slightly higher percentage of comments related to teacher feedback, meaning that learners could have paid slightly more attention to the feedback in these episodes.

Attention has been claimed to be necessary (but not necessarily sufficient) for learning to happen (Schmidt, 1990, 1993, 2001). The slightly more attention in elicit episodes should have led to more correct answers in tests. This proved true in both the immediate test and delayed test. In both tests, learners provided more correct answers for test items developed from elicit episodes than those from provide episodes (75.0% versus 63.7% in the immediate test and 70.0% versus 51.9% in the delayed test) although the difference between the two is not statistically significant (p value is .269 in the immediate test and .118 in the delayed test).

The fourth characteristic examined is linguistic focus. It describes what language aspect is targeted at in a feedback episode. Adjusted standardized residuals indicate that

grammar and spelling episodes resulted in significantly more uptake (ASR values is 2.2 and 2.5 respectively) while vocabulary episodes resulted in a significantly higher level of no opportunity (ASR=3.7). As mentioned earlier, grammar practice activities were included in the current data set. Since grammar was the focus of these activities, teachers might give learners more opportunities to respond to their feedback, thus leading to more uptake. As for spelling episodes, most of them (78.9%) were initiated by a learner with an inquiry such as "How do you spell...?" As illustrated in the discussion of the type of feedback episodes, learners as question initiators might be more responsive to teacher feedback. It is therefore not surprising that learners produced more uptake in spelling episodes too. Different from grammar episodes, vocabulary episodes often occurred in more meaning-based activities (e.g., talking about the differences between the banking system of the United States and that of one's own country or discussing one's first-day experience in the United Sates). The major concern therefore could be the contents of conversations, not language forms. Consequently, teachers might simply continue with a topic or start a new one once the feedback was given, thus leading to a higher level of no opportunity. Different from spelling episodes too, vocabulary episodes did not necessarily have learners as the dominant initiators of questions. Teachers could either ask a question or react to learners' mistakes. Accordingly, learners might not always be in an active position and hence would not always respond to the feedback, thus leading to a higher level of no uptake (ASR value is 1.0, not significant but higher than all the other three types of linguistic foci).

Concerning the successfulness of uptake, chi-square analysis indicates that it was significantly affected by linguistic focus (p=.000). One noticeable thing in the result is that pronunciation episodes resulted in the highest level of successful uptake (ASR=8.8). This could be explained by the nature of teachers' feedback. It was found that in an overwhelming majority of pronunciation episodes (96.1%), teachers provided the right answer when a problem was detected. With teachers' modeling, learners could easily repeat it and correct themselves. In episodes with vocabulary, grammar, or spelling as the linguistic focus, however, there were a variety of response types and learners sometimes could not simply repeat what teachers said. Another noticeable thing in the result is that inconclusive uptake constitutes 56.3% of learner responses in episodes with spelling as the linguistic focus. This could be explained by the fact that in 52.6% of spelling episodes learners started to write instead of talking once the teacher started to give the spelling of a word, which was coded as inconclusive. Still another noticeable thing in the result is that learners simply acknowledged teachers' feedback in a good percentage of vocabulary episodes (45.4%). A look at these episodes reveals that many of them (54.5%) concern word meanings. As discussed earlier, on such occasions there was not so much of a need for learners to incorporate the target feature right away as long as the task at hand could continue. Consequently, learners simply acknowledged the feedback.

As mentioned in the discussion of source, simple acknowledgement of teacher feedback could also be an indicator of attention. The fact that learners produced the highest percentage of acknowledge in vocabulary episodes does not mean they were not paying attention to teachers' feedback. Compared with grammar and pronunciation episodes

(spelling episodes were not included because of its low expected cell count in chi-square analysis), learners made more teacher-feedback-related comments after viewing vocabulary episodes. One possibility is that since a large number of vocabulary episodes concerned word meanings, it was crucial for learners to pay attention to teachers' feedback and to understand the feedback in order for an interaction to continue. In pronunciation and grammar episodes, however, prolonging a short vowel (e.g., kid pronounced as /ki:d/) in a particular context or using the wrong tense of a verb which was modified by an adverbial phrase (e.g., I go to Walmart yesterday morning) might not cause comprehension problems. Learners therefore might not pay as much attention to the feedback.

The next thing that was looked at is test results. Chi-square analysis indicates that the linguistic focus of feedback episodes did not significantly affect the results of the immediate test. Descriptive statistics, however, shows that the percentages of correct answers for vocabulary and grammar items (68.8% and 67.0% respectively) are both about 20.0% higher than pronunciation items (Spelling was again not included because of its low expected cell counts in the chi-square analysis). This result seems somewhat surprising given the fact that learners produced the highest percentage of successful uptake in pronunciation episodes. One possible explanation for this is that the big majority of pronunciation problems (89.5%) did not cause communication breakdowns. Learners' attention to the feedback about these problems was therefore not so intense. Consequently, their memory of the correct pronunciation did not last long. Vocabulary problems, again as shown earlier, often involved word meanings and might interrupt

classroom interactions. When there was a communication breakdown, learners might pay more attention to the problem and retain the correct language feature longer. As for grammar episodes, again because the big majority of classes observed (seven out of nine) incorporated some grammar components in their lessons and there were grammar practice activities, learners could have paid more attention to the focus grammar rules and hence did better with grammar tests too.

Different from the immediate test, chi-square analysis shows that the results of the delayed test were significantly affected by the linguistic focus of feedback episodes (p=.013). One thing that stands out is that although the percentage of correct answers for vocabulary and grammar items is virtually the same in the immediate test, the percentage of vocabulary items dropped by 22% in the delayed test while that of grammar items only dropped by 2.8%. This could be because while vocabulary feedback was more extensive, spread over the lessons as the classroom interactions required, and might not come back again afterwards, some grammar rules which were the focus of grammar lessons might have been reviewed later in a new lesson. Learners therefore could have received more feedback about grammar problems and hence were able to remember the correct forms for a longer time. Another thing that stands out in the results of the delayed test is that learners provided the lowest percentage of correct answers for spelling items. This could be explained by the fact that in all the spelling episodes, teachers spelt a word letter by letter. It was probably hard for learners to remember these discrete letters two weeks later.

The fifth characteristic examined is directness. It describes how explicit teacher feedback is. Statistical analysis indicates that indirect episodes led to significantly more uptake (ASR=3.9) and successful uptake (ASR=11.5). Direct episodes, on the other hand, resulted in significantly more no uptake (ASR=4.3) and more acknowledge (ASR=7.5) as well as inconclusive feedback (ASR=6.5). This result might have to do with the type of feedback in different episodes. In indirect episodes, two major techniques were used: recasts where teachers implicitly provided the correct form and elicit where teachers gave linguistic and/or nonlinguistic cues to help learners figure out the correct answers themselves. Elicit, as illustrated in the discussion of response type, led to a high percentage of uptake and successful uptake. As for recasts, different from Lyster and Ranta's (1997) finding that they were ineffective in eliciting learner repair (successful uptake), learners in the current study were able to correct themselves 62.5% of the time when given the opportunity to do so, thus leading to a relatively higher level of uptake and successful uptake. Now that both the two major components of indirect episodes led to more uptake and successful uptake, it is not surprising that indirect episodes as a whole led to a high level of uptake and successful uptake. In direct episodes, on the other hand, teachers often provided direct language information other than doing explicit error correction. Some of the information teachers provided might not concern learners. For example, learners might not be interested in the information about a teacher-initiated question which they understood already before the feedback. They therefore might choose to either avoid responding to the feedback or simply acknowledge it or respond to it inconclusively.

About the stimulated recall interview, despite that fact that direct episodes have a much lower percentage of uptake and successful uptake, its percentage of teacher-feedback-related comments (73.5%) is 11.7% higher than that of indirect episodes (61.8%). This could be because the feedback in indirect episodes is more implicit than the feedback in explicit episodes, hence not so easy for learners to notice.

Concerning the test results, chi-square analysis indicates that they were significantly related to the directness of feedback episodes in the immediate test but not in the delayed test (p value is .023 and .484 respectively). Descriptive statistics, however, shows that the percentage of correct answers is higher for direct episodes in both tests (15.4% higher in the immediate test and 5.1% higher in the delayed test). Given the crucial role of attention in learning, this result is reasonable now that learners made significantly more teacher-feedback-related comments in the stimulated recall interviews after viewing direct episodes.

The sixth characteristic examined is complexity. It describes how many feedback moves there are in an episode. Chi-square analysis indicates that the complexity of feedback episodes had a statistically significant effect on all the five dependent variables examined: the occurrence of uptake, the successfulness of uptake, the rate of teacher-feedback-related comments, the immediate test, and the delayed test (p value ranges from .000 to .049).

In terms of the occurrence of uptake, complex feedback led to 12.4% more uptake while simple episodes led to 11.9% more no opportunity. Now that no opportunity is often not a choice by learners (e.g., topic continuation by the teacher), the less frequent occurrence of uptake in association with simple episodes does not indicate that learners were less willing to respond to teachers in these episodes. What it could mean is that in simple episodes teachers devoted less time to the linguistic form under discussion and thus allowed fewer opportunities for learners to respond to them.

As for the successfulness of uptake, the percentage of successful uptake is 11.1% higher in complex episodes than in simple episodes (49.2% and 38.1% respectively). This means learners incorporated the correct language form in their language production more often when there were multiple feedback moves in an episode. A possible explanation for this is that multiple feedback moves constitute more input, thus enabling learners to understand the target feature better and to use it correctly.

Concerning the stimulated recall comments, learners talked significantly more about teacher feedback after viewing complex episodes (ASR=2.0), with a percentage of teacher-feedback-related comments 12.6% higher than that of simple episodes (74.7% and 62.1% respectively). This could be because the multiple feedback moves had helped to draw more attention to the feedback from learners.

Regarding the test results, learners provided significantly more correct answers for items developed from complex episodes in both the immediate test and delayed test (ASR value

is 2.4 and 2.1 respectively). This shows that learners could have indeed paid more attention to teacher feedback in complex episodes as reflected in the higher percentage of teacher-feedback-related comments associated with these episodes.

The last characteristic examined is emphasis. It is the combination of directness and complexity. Heavy episodes are either direct or complex or both while light episodes are both indirect and simple. Emphasis was found to be significantly related to the occurrence of uptake, the successfulness of uptake, teacher-feedback-related comments, and the results of the immediate test (p value ranges from .000 to .014).

In terms of the occurrence of uptake, heavy episodes led to slightly more uptake than light episodes (65.8% and 61.0% respectively). Although not statistically significant (ASR=1.3), this result still could mean that learners were more responsive in heavy episodes. Light episodes, on the other hand, led to a significantly higher level of no opportunity (ASR=2.9). Now that the feedback in light episodes is both indirect and simple, this could be because many times teachers might not want to draw overtly heavy attention to the target features in these episodes and purposefully left no opportunities for learners to respond to them.

In contrast to the occurrence of uptake, results from the successfulness of uptake indicate that learners produced significantly more successful uptake in light episodes than in heavy episodes (ASR value is 4.2 against -4.2). This could be because light episodes only contained one teacher feedback move; it was therefore easy for learners to incorporate the

target feature in their own utterance. In heavy episodes, by contrast, there are a big percentage of episodes which involved multiple feedback moves (68.8%). Language forms which require multiple feedback moves might be more complex and more difficult to take in. Learners therefore might choose to simply acknowledge teachers' feedback or to respond to it inconclusively. This could be the reason why learners produced 7.0% more acknowledge and significantly more inconclusive uptake (ASR=3.9) in heavy episodes.

Given that learners produced significantly more successful uptake in light episodes, one might expect that they would talk more about teacher feedback after viewing light episodes and perform better with test items developed from light episodes. Analysis of the stimulated recall comments and test results, however, reveals that this is not true. In actuality, heavy episodes resulted in 19.1% more teacher-feedback-related comments than light episodes (72.9% and 53.8% respectively). This could be because the multiple feedback moves or/and the directness of the feedback in heavy episodes made the target features more pronounced than those treated simply and indirectly in light episodes. It proves once again that learners might pay attention to teacher feedback when their responses to the feedback are simple acknowledgement or inconclusive uptake.

Concerning the test results, chi-square analysis shows that emphasis significantly affected the results of the immediate test but not the results of the delayed test. In the immediate test, learners performed significantly better with items developed from heavy episodes than those developed from light episodes (ASR value is 4.0 against -4.0). In the delayed

test, although there is no statistically significant difference between the two kinds of test times, learners did get 13.6% more correct answers for items developed from heavy episodes. The fact that learners performed better with test items developed from heavy episodes in both tests goes well with the fact that learners talked more about teacher feedback after viewing heavy episodes. It lends support to the assumption that more attention leads to more learning.

4.8 Summary

In summary, the results reported in this chapter suggest that both the occurrence and successfulness of uptake were significantly affected by all the seven characteristics examined. Code-related, elicit, indirect, and complex episodes led to both a higher level of uptake and a higher level of successful uptake. About the other three characteristics, student-initiated episodes, spelling episodes, and heavy episodes resulted in more uptake while reactive episodes, pronunciation episodes, and light episodes resulted in more successful uptake. Learners' stimulated recall comments were significantly affected by four characteristics: type, directness, complexity, and emphasis. Student-initiated, direct, complex, and heavy episodes are associated with a higher rate of teacher-feedback-related comments. The results of the immediate test were also significantly affected by these same four characteristics in the same way, with the same four categories resulting in more correct answers. The results of the delayed test were only significantly affected by linguistic focus and directness. Learners performed better with test items developed from grammar and complex episodes. These results taken as whole show that the answer

to the first research question is: yes, the characteristics of teacher feedback episodes affect learners' noticing and learning to a certain degree.

CHAPTER 5

NONLINGUISTIC CUES, THE NOTICING AND THE EFFECT OF TEACHER FEEDBACK

5.1 Introduction

This chapter addresses the second research question: Do the nonlinguistic cues in teacher feedback affect learners' noticing and learning? After presenting the frequency of nonlinguistic cues in teacher feedback and the distribution of their subcategories, I report and discuss the relationship between nonlinguistic cues and learners' responses to teachers' feedback, learners' comments in the stimulated recall interviews, and results from both the immediate test and delayed test. For purposes of this analysis, "nonlinguistic" includes paralinguistic cues such as word stress and intonation, and extralinguistic cues such as gestures and head movements.

5.2 Occurrence and distribution of nonlinguistic cues

This section reports the occurrence of general paralinguistic cues and general extralinguistic cues as well as the distribution of their subtypes. Table 50 shows that teachers used paralinguistic cues in 22.3% of a total of 1434 feedback episodes. The percentage for episodes with extralinguistic cues is 40.7%. Table 51 shows that among a total of 320 paralinguistic cues, 52.5% are word stress, 10.9% are rising intonation, 14.4% are eliciting stop, and 11.9% are dragging voice. Mimicking sound (3.1%) and combination (7.2%) take the rest 10.3%. Concerning the type of extralinguistic cues, among a total of 548 extralinguistic cues, 58.4% are gestures, 15.1% are head movements,

Table 50 Occurrence of nonlinguistic cues (N=1434)

	Number	Percentage
Paralinguistic cues		
Present	320	22.3%
Absent	1114	77.7%
Extralinguistic cues		
Present	584	40.7%
Absent	850	59.3%

Table 51 Distribution of different types of nonlinguistic cues

	Number	Percentage
Type of paralinguistic cues		
Word stress	168	52.5%
Rising intonation	35	10.9%
Eliciting stop	46	14.4%
Dragging voice	38	11.9%
Mimicking sound	10	3.1%
Combination	23	7.2%
Total	320	100%
Type of extralinguistic cues		
Gestures	341	58.4%
Whole body acting	29	5.0%
Head movements	88	15.1%
Facial expressions	12	2.1%
Combination	114	19.5%
Total	584	100%
Type of gestures		
Iconics	139	40.8%
Metaphorics	38	11.1%
Deictics	99	29.0%
Beats	24	7.0%
Combination	41	12.0%
Total	341	100%

and 19.5% are combinations. Whole body acting (5.0%) and facial expressions (2.1%) take the rest 7.1%. The distribution of different types of gestures is presented in the same table to save space. Among a total of 341 gestures, 40.8% are iconic, 11.1% are metaphoric, 29.0% are deictic, and 12.0% are combinations. With a percentage of 7.0%, beats occurred most infrequently.

5.3 Nonlinguistic cues and learner uptake

This section reports the relationship between the nonlinguistic cues in teacher feedback and learner uptake. The occurrence and successfulness of uptake are presented together according to each type of nonlinguistic cues and their subcategories in the order of general paralinguistic cues, type of paralinguistic cues, general extralinguistic cues, type of extralinguistic cues, and type of gestures.

General paralinguistic cues

Tables 52 and 53 illustrate the occurrence and successfulness of uptake according to the presence of general paralinguistic cues. Table 52 shows that learners responded to teacher feedback in 73.1% of episodes with paralinguistic cues and 62.9% of episodes without paralinguistic cues. Adjusted standardized residuals in the table reveal that uptake was significantly more frequent in episodes with paralinguistic cues while no opportunity was significantly more frequent in episodes without paralinguistic cues. There is no significant difference between the two in terms of no uptake. Table 53 shows that when uptake did occur, 53.4% was successful in episodes with paralinguistic cues and 42.5% was successful in episodes without paralinguistic cues. Adjusted standardized residuals in

the table reveal that episodes with paralinguistic cues led to significantly more successful uptake while episodes without paralinguistic cues led to significantly more acknowledge. There is no significant difference between the two in terms of unsuccessful and inconclusive uptake. Chi-square tests indicate that both the occurrence and successfulness of uptake are significantly related to teachers' use of general paralinguistic cues (*p* value is .000 for the occurrence of uptake and .006 for the successfulness of uptake).

Table 52 General paralinguistic cues and the occurrence of uptake

	Number of episodes	Occurrence of uptake		
		Uptake	No uptake	No opportunity
Present	320	234 (73.1%)	62 (19.4%)	24 (7.5%)
Residuals (Present)	--	3.4	-1.0	-3.5
Absent	1114	701 (62.9%)	244 (21.9%)	169 (15.2%)

$X^2(2, n=1434)=15.578, p=.000$

Table 53 General paralinguistic cues and the successfulness of uptake

	Number of uptake	Successfulness of uptake			
		Successful	Unsuccessful	Acknowledge	Inconclusive
Present	234	125 (53.4%)	4 (1.7%)	66 (28.2%)	39 (16.7%)
Residuals (Present)	--	2.9	-.1	-3.4	.7
Absent	701	298 (42.5%)	13 (1.8%)	286 (40.8%)	104 (14.8%)

$X^2(3, n=935)=12.410, p=.006$

Type of paralinguistic cues

Tables 54, 55, and 56 illustrate the occurrence and successfulness of uptake according to the type of paralinguistic cues. Table 54 shows that learners responded to teacher feedback in 70.8% of episodes with word stress, 80.0% of episodes with rising intonation, 73.9% of episodes with eliciting stop, and 76.3% of episodes with dragging voice. Table 55 shows that when uptake did occur, 44.9% was successful in episodes with word stress,

74.1% was successful in episodes with rising intonation, 63.6% was successful in episodes with eliciting stop, and 62.1% was successful in episodes with dragging voice. Adjusted standardized residuals in Table 56 reveal that episodes with rising intonation led to significantly more successful uptake while episodes with word stress led to significantly more acknowledge. There is no significant difference among the four in terms of inconclusive uptake. Chi-square tests indicate that while the occurrence of uptake is not significantly related to the type of paralinguistic cues, the successfulness of uptake is (*p* value is .619 for the occurrence of uptake and .050 for the successfulness of uptake).

Table 54 Type of paralinguistic cues and the occurrence of uptake

	Number of episodes	Occurrence of uptake		
		Uptake	No uptake	No opportunity
Word stress	168	119 (70.8%)	37 (22.0%)	12 (7.1%)
Rising intonation	35	28 (80.0%)	4 (11.4%)	3 (8.6%)
Eliciting Stop	46	34 (73.9%)	11 (23.9%)	1 (2.2%)
Dragging voice	38	29 (76.3%)	6 (15.8%)	3 (7.9%)

$X^2(6, n=287)=4.429$, $p=.619$ (Mimicking sound and combination were excluded due to low expected cell counts)

Table 55 Type of paralinguistic cues and the successfulness of uptake

	Number of uptake	Successfulness of uptake		
		Successful	Acknowledge	Inconclusive
Word stress	118	53 (44.9%)	45 (38.1%)	20 (16.9%)
Rising intonation	27	20 (74.1%)	4 (14.8%)	3 (11.1%)
Eliciting Stop	33	21 (63.6%)	8 (24.2%)	4 (12.1%)
Dragging voice	29	18 (62.1%)	5 (17.2%)	6 (20.7%)

$X^2(6, n=207)=12.536$, $p=.050$ (Mimicking sound, combination, partially successful uptake, and unsuccessful uptake were excluded due to low expected cell counts)

Table 56 Type of paralinguistic cues and the successfulness of uptake residuals

	Successful	Acknowledge	Inconclusive
Word stress	-3.1	3.0	.5
Rising intonation	2.2	-1.8	-.7
Eliciting Stop	1.2	-.8	-.7
Dragging voice	.9	-1.6	.8

General extralinguistic cues

Tables 57 and 58 illustrate the occurrence and successfulness of uptake according to the presence of general extralinguistic cues. Table 57 shows that learners responded to teacher feedback in 65.9% of episodes with extralinguistic cues and 64.7% of episodes without extralinguistic cues. Adjusted standardized residuals in the table reveal that no opportunity was significantly more frequent in episodes without extralinguistic cues than those with extralinguistic cues but there is no significant difference between the two in terms of uptake and no uptake. Table 58 shows that when uptake did occur, 38.4% was successful in episodes with extralinguistic cues and 50.0% was successful in episodes without extralinguistic cues. Adjusted standardized residuals in the table reveal that episodes without extralinguistic cues led to significantly more successful uptake while episodes with extralinguistic cues led to significantly more acknowledge. There is no significant difference between the two in terms of unsuccessful and inconclusive uptake. Chi-square tests indicate that both the occurrence and successfulness of uptake are significantly related to teachers' use of general extralinguistic cues (p value is .047 for uptake and .005 for the successfulness of uptake)

Table 57 General extralinguistic cues and the occurrence of uptake

	Number of episodes	Occurrence of uptake		
		Uptake	No uptake	No opportunity
Present	584	385 (65.9%)	135 (23.1%)	64 (11.0%)
Residuals (present)	--	.5	1.4	-2.3
Absent	850	550 (64.7%)	171 (20.1%)	129 (15.2%)

$X^2(2, n=1434)=6.113, p=.047$

Table 58 General extralinguistic cues and the successfulness of uptake

	Number of uptake	Successfulness of uptake			
		Successful	Unsuccessful	Acknowledge	Inconclusive
Present	385	148 (38.4%)	7 (0.8%)	166 (43.1%)	64 (16.6%)
Residuals (Present)	--	-3.5	.0	2.9	.9
Absent	550	275 (50.0%)	10 (1.8%)	186 (33.8%)	79 (14.4%)

$X^2(3, n=935)=12.645, p=.005$

Type of extralinguistic cues

Tables 59, 60, and 61 illustrate the occurrence and successfulness of uptake according to the type of extralinguistic cues. Table 59 shows that learners responded to teacher feedback in 61.6% of episodes with gestures, 72.4% of episodes with whole body acting, 83.0% of episodes with head movements, 66.7% of episodes with facial expressions, and 64.0% of episodes with the combination of different types of extralinguistic cues. Adjusted standardized residuals in Table 60 reveal that uptake was significantly more frequent in episodes with head movements, no uptake was significantly more frequent in episodes with the combination of different types of extralinguistic cues, and no opportunity was significantly more frequent in episodes with gestures. Table 61 shows that when uptake did occur, 33.3% was successful in episodes with gestures, 47.7% was successful in episodes with whole body acting, 51.4% was successful in episodes with head movements, and 37.1% was successful in episodes with the combination of different

types of extralinguistic cues. Chi-square tests indicate that while the occurrence of uptake is significantly related to the type of extralinguistic cues, the successfulness of uptake is not (p value is .000 for the occurrence of uptake and .171 for the successfulness of uptake).

Table 59 Type of extralinguistic cues and the occurrence of uptake

	Number of episodes	Occurrence of uptake		
		Uptake	No uptake	No opportunity
Gestures	341	210 (61.6%)	77 (22.6%)	54 (15.8%)
Whole body acting	29	21 (72.4%)	8 (27.6%)	0 (0%)
Head movements	88	73 (83.0%)	12 (13.6%)	3 (3.4%)
Facial expressions	12	8 (66.7%)	3 (25.0%)	1 (8.3%)
Combination	114	73 (64.0%)	35 (30.7%)	6 (5.3%)

$X^2(8, n=584)=30.270, p=.000$

Table 60 Type of extralinguistic cues and the occurrence of uptake residuals

	Uptake	No uptake	No opportunity
Gestures	-2.6	-.4	4.5
Whole body acting	.8	.6	-1.9
Head movements	3.7	-2.3	-2.5
Facial expressions	.1	.2	-.3
Combination	-.5	2.1	-2.2

Table 61 Type of extralinguistic cues and the successfulness of uptake

	Number of uptake	Successfulness of uptake		
		Successful	Acknowledge	Inconclusive
Gestures	207	69 (33.3%)	97 (46.9%)	41 (19.8%)
Whole body acting	21	10 (47.7%)	9 (42.9%)	2 (9.5%)
Head movements	72	37 (51.4%)	25 (34.7%)	10 (13.9%)
Combination	70	26 (37.1%)	33 (47.1%)	11 (15.7%)

$X^2(6, n=370)=9.049, p=.171$ (Facial expressions, partially successful uptake, and unsuccessful uptake were excluded due to low expected cell counts)

Type of gestures

Tables 62 and 63 illustrate the occurrence and successfulness of uptake according to the type of gestures. Table 62 shows that learners responded to teacher feedback in 59.0% of episodes with iconic gestures, 68.4% of episodes with metaphoric gestures, 64.6% of episodes with deictic gestures, 66.7% of episodes with beats, and 53.7% of episodes with the combination of different types of gestures. Table 63 shows that when uptake did occur, 28.4% was successful in episodes with iconic gestures, 34.6% was successful in episodes with metaphoric gestures, 28.6% was successful in episodes with deictic gestures, 50.0% was successful in episodes with beats, and 52.4% was successful in episodes with the combination of different types of gestures. Chi-square tests indicate

Table 62 Type of gestures and the occurrence of uptake

	Number of episodes	Occurrence of uptake		
		Uptake	No uptake	No opportunity
Iconics	139	82 (59.0%)	34 (24.5%)	23 (16.5%)
Metaphorics	38	26 (68.4%)	5 (13.2%)	7 (18.4%)
Deictics	99	64 (64.6%)	21 (21.2%)	14 (14.1%)
Beats	24	16 (66.7%)	6 (25.0%)	2 (8.3%)
Combination	41	22 (53.7%)	11 (26.8%)	8 (19.5%)

$X^2(8, n=341)=4.881, p=.770$

Table 63 Type of gestures and the successfulness of uptake

	Number of uptake	Successfulness of uptake		
		Successful	Acknowledge	Inconclusive
Iconics	81	23 (28.4%)	42 (51.9%)	16 (19.8%)
Metaphorics	26	9 (34.6%)	12 (46.2%)	5 (19.2%)
Deictics	63	18 (28.6%)	28 (44.4%)	17 (27.0%)
Beats	16	8 (50.0%)	7 (43.8%)	1 (6.3%)
Combination	21	11 (52.4%)	8 (38.1%)	2 (9.5%)

$X^2(8, n=207)=9.791, p=.280$ (Partially successful uptake and unsuccessful uptake were excluded due to low expected cell counts)

that neither the occurrence nor the successfulness of uptake is significantly related to the type of gestures (p value is .770 for the occurrence of uptake and .280 for the successfulness of uptake).

To sum up, the descriptive and inferential analysis in this section reveals that among the five factors examined, three of them showed a statistically significant effect on the occurrence of uptake: general paralinguistic cues, general extralinguistic cues, and the type of extralinguistic cues ($p<.05$). Adjusted standardized residuals reveal that learners responded to teacher feedback significantly more often in episodes where teachers used general paralinguistic cues and where they used head movements (ASR>2.0). Regarding the successfulness of uptake, among the five factors examined, general paralinguistic cues, general extralinguistic cues, and the type of paralinguistic cues exerted a statistically significant effect on it. Adjusted standardized residuals indicate that learners produced significantly more successful uptake in episodes where teachers used general paralinguistic cues and where they used rising intonation.

5.4 Nonlinguistic cues and recall comments

This section reports the relationship between the nonlinguistic cues in teacher feedback and learners' comments in the stimulated recall interviews. The frequency of comments related to teacher feedback is presented in relation to each type of nonlinguistic cues and their subcategories in the order of general paralinguistic cues, type of paralinguistic cues, general extralinguistic cues, type of extralinguistic cues, and type of gestures.

General paralinguistic cues

Table 64 shows the rate of teacher-feedback-related comments in relation to the presence of general paralinguistic cues. A total of 77.1% of recall comments are related to teacher feedback for episodes with paralinguistic cues and 70.3% are related to teacher feedback for episodes without paralinguistic cues. Chi-square test results suggest that the rate of teacher-feedback-related comments was not significantly affected by the presence of general paralinguistic cues (p=.493).

Table 64 General paralinguistic cues and teacher-feedback-related comments

	Total number of comments	Teacher-feedback-related	Non-teacher-feedback-related	Other
Present	70	54 (77.1%)	7 (10.0%)	9 (12.9%)
Absent	209	147 (70.3%)	23 (11.0%)	39 (18.6%)

$X^2(2, n=279)=1.413, p=.493$

Type of paralinguistic cues

Table 65 shows the rate of teacher-feedback-related comments in relation to the type of paralinguistic cues. A total of 88.5% of recall comments are related to teacher feedback for episodes with word stress, 88.9% are related to teacher feedback for episodes with rising intonation, 92.9% are related to teacher feedback for episodes with eliciting stop, 85.7% are related to teacher feedback for episodes with dragging voice, 75.0% are related to teacher feedback for episodes with mimicking sound, and 100% are related to teacher feedback for episodes with the combination of different types of paralinguistic cues. There is no case for the categories of "inconclusive" and "other". Chi-square analysis was not performed due to low expected cell counts.

Table 65 Type of paralinguistic cues and teacher-feedback-related comments

	Total number of comments	Teacher-feedback-related	Non-teacher-feedback-related
Word stress	26	23 (88.5%)	3 (11.5%)
Rising intonation	9	8 (88.9%)	1 (11.1%)
Eliciting Stop	14	13 (92.9%)	1 (7.1%)
Dragging voice	7	6 (85.7%)	1 (14.3%)
Mimicking sound	4	3 (75.0%)	1 (25.0%)
Combination	1	1 (100%)	0 (0%)

General extralinguistic cues

Table 66 shows the rate of teacher-feedback-related comments in relation to the presence of general extralinguistic cues. A total of 77.0% of recall comments are related to teacher feedback for episodes with extralinguistic cues and 62.5%% are related to teacher feedback for episodes without extralinguistic cues. Adjusted standardized residuals indicate that learners talked significantly more about teacher feedback after viewing episodes with extralinguistic cues. Chi-square test results suggest that the rate of teacher-feedback-related comments was significantly affected by the presence of general extralinguistic cues (p=.011).

Table 66 General extralinguistic cues and teacher-feedback-related comments

	Total number of comments	Teacher-feedback-related	Non-teacher-feedback-related	Other
Present	183	141 (77.0%)	13 (7.1%)	29 (15.8%)
Residuals (Present)	--	2.6	-2.7	-.8
Absent	96	60 (62.5%)	17 (17.7%)	19 (19.8%)

$X^2(2, n=279)=9.005, p=.011$

Type of extralinguistic cues

Table 67 shows the rate of teacher-feedback-related comments in relation to the type of extralinguistic cues. A total of 76.4% of recall comments are related to teacher feedback for episodes with gestures, 60.0% are related to teacher feedback for episodes with whole body acting, 80.0% are related to teacher feedback for episodes with head movements, 50.0% are related to teacher feedback for episodes with facial expressions, and 80.0% are related to teacher feedback for episodes with the combination of different types of extralinguistic cues. Chi-square analysis was not performed due to low expected cell counts.

Table 67 Type of extralinguistic cues and teacher-feedback-related comments

	Total number of comments	Teacher-feedback-related	Non-teacher-feedback-related	Inconclusive	Other
Gestures	106	81 (76.4%)	8 (7.5%)	2 (1.9%)	15 (14.2%)
Whole body acting	5	3 (60.0%)	0 (0%)	0 (0%)	2 (40.0%)
Head movements	20	16 (80.0%)	0 (0%)	0 (0%)	4 (20%)
Facial expressions	2	1 (50.0%)	0 (0%)	0 (0%)	1 (50.0%)
Combination	50	40 (80.0%)	5 (10.0%)	1 (2.0%)	4 (8.0%)

Type of gestures

Table 68 shows the rate of teacher-feedback-related comments in relation to the type of gestures. A total of 81.6% of recall comments are related to teacher feedback for episodes with iconic gestures, 62.5% are related to teacher feedback for episodes with metaphoric gestures, 65.4% are related to teacher feedback for episodes with deictic gestures, 80.0% are related to teacher feedback for episodes with beats, and 93.8% are related to teacher

feedback for episodes with the combination of different types of gestures. Chi-square analysis was not performed due to low expected cell counts.

Table 68 Type of gestures and teacher-feedback-related comments

	Total number of comments	Teacher-feedback-related	Non-teacher-feedback-related	Other
Iconics	49	40 (81.6%)	3 (6.1%)	6 (12.2%)
Metaphorics	8	5 (62.5%)	0 (0%)	3 (37.5%)
Deictics	26	17 (65.4%)	4 (15.4%)	5 (19.2%)
Beats	5	4 (80.0%)	0 (0%)	1 (20.0%)
Combination	16	15 (93.8%)	1 (6.2%)	0 (0%)

To sum up, in terms of nonlinguistic cues and stimulated recall comments, chi-square analysis was only performed for the presence of general paralinguistic cues and general extralinguistic cues. Between these two, only extralinguistic cues had a significant effect on the rate of teacher-feedback-related comments. That is, learners talked significantly more about teacher feedback after viewing episodes with extralinguistic cues. It was not clear whether the type of paralinguistic cues, the type of extralinguistic cues, and the type of gestures had significantly affected the rate of teacher-feedback-related comments because low expected cell counts made it impossible to perform valid chi-square tests.

5.5 Nonlinguistic cues and test results

This section reports the relationship between the nonlinguistic cues in teacher feedback and test results. Results from both the immediate test and delayed test are presented together according to each type of nonlinguistic cues and their subcategories in the order of general paralinguistic cues, type of paralinguistic cues, general extralinguistic cues, type of extralinguistic cues, and type of gestures.

General paralinguistic cues

Table 69 is an illustration of test results according to the presence of general paralinguistic cues. In the immediate test, the percentage of correct answers is 68.9% for episodes with paralinguistic cues and 63.1% for episodes without paralinguistic cues. In the delayed test, the percentage of correct answers for the two is 56.3% and 52.0% respectively. Chi-square analysis indicates that the presence of general paralinguistic cues did not significantly affect the results of either test (*p* value is .374 and .517 respectively)

Table 69 General paralinguistic cues and test results

	Immediate test			Delayed test		
	Number of test items	Correct	Incorrect	Number of test items	Correct	Incorrect
Present	74	51 (68.9%)	23 (31.1%)	80	45 (56.3%)	35 (43.8%)
Absent	198	125 (63.1%)	73 (36.9%)	202	105 (52.0%)	97 (48.1%)
	$X^2(1, n=272)=.790, p=.374$			$X^2(1, n=282)=420, p=.517$		

Type of paralinguistic cues

Table 70 is an illustration of test results according to the type of paralinguistic cues. In the immediate test, the percentage of correct answers is 76.0% for episodes with word stress, 66.7% for episodes with eliciting stop, 62.5 for episodes with dragging voice, and 87.5% for episodes with the combination of different types of paralinguistic cues. In the delayed test, the percentage of correct answers for the four is 56.1%, 100%, 40.0%, and 83.3% respectively. The percentage is 28.6% for episodes with rising intonation and 66.7% for episodes with mimicking sound. Chi-square analysis indicates that the type of

paralinguistic cues did not significantly affect the results of the immediate test (p=.588). No chi-square test was performed for the delayed test due to low expected cell counts.

Table 70 Type of paralinguistic cues and test results

	Immediate test			Delayed test		
	Number of test items	Correct	Incorrect	Number of test items	Correct	Incorrect
Word Stress	25	19 (76.0%)	6 (24.0%)	57	32 (56.1%)	25 (43.9%)
Eliciting stop	30	20 (66.7%)	10 (33.3%)	2	2 (100%)	0 (0%)
Dragging voice	8	5 (62.5%)	3 (37.5%)	5	2 (40.0%)	3 (60.0%)
Combination	8	7 (87.5%)	1 (12.5%)	6	5 (83.3%)	1 (16.7%)
Rising intonation	Excluded due to low expected cell counts			7	2 (28.6%)	5 (71.4%)
Mimicking sound	Excluded due to low expected cell counts			3	2 (66.7%)	1 (33.3%)
	$X^2(3, n=71)=1.925, p=.588$					

General extralinguistic cues

Table 71 is an illustration of test results according to the presence of general extralinguistic cues. In the immediate test, the percentage of correct answers is 73.0% for episodes with extralinguistic cues and 55.7% for episodes without extralinguistic cues. In the delayed test, the percentage of correct answers for the two is 60.0% and 47.1% respectively. Adjusted standardized residuals suggest that learners provided significantly more correct answers for test items developed from episodes with extralinguistic cues in both the immediate test and delayed test. Chi-square analysis indicates that the presence of general extralinguistic cues significantly affected the results of both tests (p value is .003 and .023 respectively).

Table 71 General extralinguistic cues and test results

	Immediate test				Delayed test			
	Number of test items	Correct	Residuals (Correct)	In-correct	Number of test items	Correct	Residuals (Correct)	In-correct
Present	141	103 (73.0%)	3.0	38 (27.0%)	127	77 (60.6%)	2.3	50 (39.4%)
Absent	131	73 (55.7%)	-3.0	58 (44.3%)	155	73 (47.1%)	-2.3	82 (52.9%)
	$X^2(1, n=272)=8.925, p=.003$				$X^2(1, n=282)=5.135, p=.023$			

Type of extralinguistic cues

Table 72 is an illustration of test results according to the type of extralinguistic cues. In the immediate test, the percentage of correct answers is 73.6% for episodes with gestures, 63.7% for episodes with head movements, and 71.1% for episodes with the combination of different types of extralinguistic cues. In the delayed test, the percentage of correct answers for the three is 54.3%, 52.4%, and 75.8% respectively. Chi-square analysis indicates that the type of extralinguistic cues did not significantly affect the results of either test (*p* value is .777 and .088 respectively).

Table 72 Type of extralinguistic cues and test results

	Immediate test			Delayed test		
	Number of test items	Correct	Incorrect	Number of test items	Correct	Incorrect
Gestures	87	64 (73.6%)	23 (26.4%)	70	38 (54.3%)	32 (45.7%)
Head movements	11	7 (63.7%)	4 (36.3%)	21	11 (52.4%)	10 (47.6%)
Combination	38	27 (71.1%)	11 (28.9)	33	25 (75.8%)	8 (24.2%)
	$X^2(2, n=136)=.504, p=.777$			$X^2(2, n=124)=4.856, p=.088$		

(Whole body acting and facial expressions were excluded from both tests due to low expected cell counts)

Type of gesture

Table 73 is an illustration of test results according to the type of gestures. In the immediate test, the percentage of correct answers is 65.7% for episodes with iconic gestures, 90.0% for episodes with metaphoric gestures, 84.0% for episodes with deictic gestures, and 40.0% for episodes with beats. Adjusted standardized residuals suggest that learners provided more (although not significantly) correct answers for items developed from metaphoric and deictic gestures and significantly less correct answers for test items developed from episodes with beats. In the delayed test, the percentages of correct answers for iconic, metaphoric, and deictic gestures are 52.0%, 66.7%, and 55.0% respectively. Chi-square analysis indicates that the type of gestures significantly affected the results of the immediate test (p=.029) but not the results of the delayed test (p=.748).

Table 73 Type of gestures and test results

	Immediate test				Delayed test		
	Number of test items	Correct	Residuals (Correct)	Incorrect	Number of test items	Correct	Incorrect
Iconics	35	23 (65.7%)	-1.0	12 (34.3%)	25	13 (52.0%)	12 (48.0%)
Metaphorics	10	9 (90.0%)	1.4	1 (10.0%)	9	6 (66.7%)	3 (33.3%)
Deictics	25	21 (84.0%)	1.7	4 (16.0%)	20	11 (55.0%)	9 (45.0%)
Beats	10	4 (40.0%)	-2.3	6 (60.0%)	Excluded due to low expected cell counts		
	$X^2(3, n=80)=8.991, p=.029$				$X^2(2, n=54)=.581, p=.748$		

(Combination was excluded from both tests due to low expected cell counts)

To sum up, the descriptive and inferential analysis in this section reveals that general extralinguistic cues had a statistically significant effect on the results of both the

immediate test and delayed test ($p<.05$), resulting in significantly more correct answers when they were used in teacher feedback (ASR>2.0). The type of gestures also significantly affected the immediate test, with deictic gestures associated with more correct answers although the value of the adjusted standardized residual is not up to 2.0. On the other hand, no such pattern was found in the delayed test. The type of paralinguistic cues did not have any statistically significant effect on the immediate test and it is not clear whether it had an effect on the delayed test now that no chi-square analysis was performed due to low expected cell counts. The other two factors, general paralinguistic cues and the type of extralinguistic cues, did not show any statistically significant effect on either test.

5.6 Review of results

The results presented above show that the occurrence of uptake was significantly affected by the presence of general paralinguistic cues, the presence of general extralinguisticues, and the type of extralinguistic cues but not by the type of paralinguistic cues and the type of gestures. The successfulness of uptake was significantly related to the presence of general paralinguistic cues, the type of paralinguistic cues, and the presence of general extralinguistic cues but not by the type of extralinguistic cues and the type of gestures. In terms of the rate of teacher-feedback-related recall comments, while the presence of general paralinguistic cues did not have any statistically significant effect on it, the presence of general extralinguistic cues did. It is not clear whether the type of paralinguistic cues, the type of extralinguistic cues, and the type of gestures had any statistically significant bearing on the recall comments because no chi-square analysis

could be conducted with these three variables. As for the tests, results of the immediate test were significantly affected by the presence of general extralinguistic cues and the type of gestures but not by the presence of general paralinguistic cues, the type of paralinguistic cues, and the type of extralinguistic cues. The only variable that was shown to have significantly affected the results of the delayed test is again the presence of general extralinguistic cues. It is not clear whether the type of paralinguistic cues had such an effect on the delayed test because no chi-square analysis could be performed with this variable. The table below is a summary of the overall results in this section. In addition to the plus sign and minus sign, a question mark is used to indicate a statistically unknown relationship.

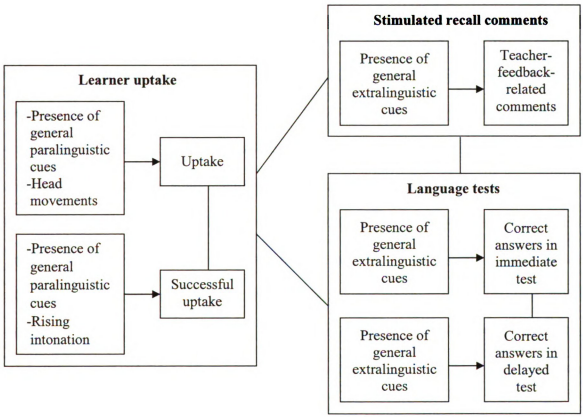
Table 74 Overall results by nonlinguistic cues

	Occurrence of uptake	Successfulness of uptake	Feedback-related comments	Immediate test	Delayed test
General paralinguistic cues	+	+	-	-	-
Type of paralinguistic cues	-	+	?	-	?
General extralinguistic cues	+	+	+	+	+
Type of extralinguistic cues	+	-	?	-	-
Type of gestures	-	-	?	+	-

Figure 6 is an overview of nonlinguistic cues and type of nonlinguistic cues that predicted uptake, successful uptake, teacher-feedback-related comments, and correct test results.

Specifically, episodes with general paralinguistic cues and head movements led to more uptake. General paralinguistic cues, together with rising intonation, also contributed to more successful uptake. By contrast, teacher-feedback-related comments and correct answers for both the immediate test and delayed test were predicted by the presence of general extralinguistic cues.

Figure 6 Overview of nonlinguistic cues predicting (successful) uptake, teacher-feedback-related comments, and correct test results



5.7 Discussion:

This section discusses the results presented in the previous sections. The occurrence and successfulness of uptake are examined together with learners' stimulated recall comments

and test results in relation to general paralinguistic cues, the type of paralinguistic cues, general extalinguistic cues, the type of extralinguistic cues, and the type of gestures.

As shown in Chapter 2, very little research has been conducted in the field of second language acquisition to systematically examine the effect of nonlinguistic cues on noticing and learning. The few studies that did make an effort to address this issue (Carpenter et al., 2006; Davies, 2006; Faraco and Kida, 2008; Loewen & Philp, 2006; Sheen, 2004; Sime, 2006) only focused on some aspects of nonlinguistic cues. Although there is some counterevidence, the general conclusion from these studies is that nonlinguistic cues have a positive effect on noticing and/or learning. Different from these studies, the present study attempted to look at the effect of both general nonlinguistic cues and specific types of nonlinguistic cues. It was found that both paralinguistic cues and extralinguistic cues significantly affected some of the dependent variables when examined at the general level. This partially confirms the findings from existing studies. At the more specific level, there are no significant differences among the different types of nonlinguistic cues in most cases. Among the six types of paralinguistic cues, only intonation had a strong positive effect on one of the five major dependent variables, the successfulness of uptake, confirming the findings of Loewen and Philp (2006) and Sheen (2006). Among the five types of extralinguistic cues, only head movements had a strong positive effect on another one of the five dependent variables, uptake. Gestures, a type of extralinguistic cues which Faraco and Kida (2008) and Sime (2006) found to help clarify verbally expressed meanings, did not stand out from the results. Below is a detailed examination of nonlinguistic cues.

General paralinguistic cues refer to any voice feature that is part of language but are not linguistic in nature. The word "general" is used because this term does not concern itself with any specific kind of cues. It was found that both the occurrence and successfulness of uptake were significantly affected by paralinguistic cues (p value is .000 and .006 respectively). Learners produced significantly more uptake and more successful uptake when paralinguistic cues were present in teacher feedback (ASR value is 3.4 and 2.9 respectively). This result lends support to Lowen and Philp's (2006) finding that prosodic features such as emphasis and intonation can increase the effectiveness of recasts. It means paralinguistic cues can help engage learners in teacher-student interactions and push them to incorporate the correct form in their language production.

Although the presence of paralinguistic cues significantly affected the occurrence and successfulness of uptake, it did not show any statistically significant effect on the rate of teacher-feedback-related comments or the results of the two tests (p>.05 in all three cases). One possible explanation is that the paralinguistic cues identified in the current data set, along with those that were not examined (e.g., speech rate and voice quality), averaged teachers' speech so that they became less pronounced to learners. A second possibility is that paralinguistic cues as subtle aspects of speech (Mehrebian, 1972) are not as noticeable as verbal words themselves. On the other hand, compared with episodes without paralinguistic cues, learners did make a slightly higher percentage of teacher-feedback-related comments after viewing episodes with paralinguistic cues (77.1% versus 70.3%) and provided a higher percentage of correct answers for test items developed

from episodes with paralinguistic cues in both tests (68.9% versus 63.1% in the immediate test and 56.3% versus 52.0% in the delayed test). The slightly higher percentages of teacher-feedback-related comments and correct answers, along with the fact that there are significantly more uptake and more successful uptake in episodes with paralinguistic cues, indicate that paralinguistic cues have the potential to help arouse learners' attention to teacher feedback and improve learning.

Unlike general paralinguistic cues, the type of paralinguistic cues is concerned with different kinds of acoustic nonlinguistic cues. Chi-square analysis indicates that there is no statistically significant difference among different types of paralinguistic cues in terms of the occurrence of uptake (p=.619). This means that the effect of different types of paralinguistic cues on the occurrence of uptake is similar to one another. In terms of the successfulness of uptake, however, learners produced significantly more successful uptake in episodes with rising intonation (ASR=2.2) and significantly less successful uptake in episodes with word stress (ASR=-3.1). In 66.7% of episodes with rising intonation, learners made an erroneous utterance and teachers repeated the problematic language form in order to prompt learners to correct themselves (e.g., I would some coffee?). Rising intonation was therefore often present in eliciting episodes, which has been shown to lead to a high level of successful uptake. Given that, it is reasonable that episodes with rising intonation led to more successful uptake. Word stress, although also used in some eliciting episodes (generally together with rising intonation), were largely (94.0%) employed when teachers were providing linguistic information to learners (e.g., You have class in 30 minutes. IN 30 minutes. Not after). As discussed in Chapter 4,

provide episodes tend to result in less successful uptake and more acknowledge because learners were in a more passive position. The frequent presence of word stress in information provision episodes therefore could have resulted in the lower level of successful uptake and higher level of acknowledge (ASR=3.0).

Another noticeable thing about the type of paralinguistic cues is eliciting stop. Like rising intonation, eliciting stop is also a technique teachers often used to elicit language information from learners. However, its level of successful uptake is lower than that of rising intonation (ASR value is 1.2 against 2.2). It is not clear exactly why this is so, but it could be due to the nature of the two types of techniques. With rising intonation, teachers expected learners to correct themselves and learners were supposedly able to get the right form with the proper amount of help because they at least had had some partial knowledge of the language form already. With eliciting stop, however, learners were expected to provide language forms or information they might not know at all. For example, one teacher wanted to elicit the word "abandon" from learners, so he gave the first part of the word aban- and then stopped for learners to give the whole word. It turned out that learners were not able to provide the rest of the word and the teacher had to provide it himself. Different from episodes with rising intonation where learners had produced an utterance already, in this case learners did not seem to have any idea of the target feature at all. Given the nature of eliciting stop, it is not surprising that it led to a lower level of successful uptake than rising intonation.

With respect to the stimulated recall interviews, no chi-square analysis was conducted due to low expected cell counts. It is therefore not clear whether there is a statistically significant relationship between the type of paralinguistic cues and teacher-feedback-related comments. The percentages of such comments, however, indicate that there is no big difference among the different types of paralinguistic cues. For the three types that have been discussed in terms of the successfulness of uptake (word stress, rising intonation, and eliciting stop), the percentage of teacher-feedback-related comments is very close (88.5%, 88.9%, and 92.9% respectively). This reinforces what was discussed previously in Chapter 4: learners' responses to teacher feedback other than successful uptake (e.g., acknowledge and inconclusive uptake) can also mean noticing.

Concerning test results, no chi-square analysis was performed for the delayed test due to low expected cell counts. It is therefore not clear whether the different types of paralinguistic cues significantly affected the test results. Chi-square analysis of the immediate test indicates that it did not. Descriptive statistics shows that the combination of different kinds of paralinguistic cues has the highest percentage of correct answers in the immediate test (87.5%). This result, as suspected, could be because more than one type of paralinguistic cues together could create a bigger effect than one single type of paralinguistic cues. This hypothesis is upheld by the still relatively high percentage of correct answers resulted from episodes with the combination of different types of paralinguistic cues in the delayed test (83.3%) although this percentage is lower than that of eliciting stop. Moreover, there are only two items for eliciting stop, both of which learners answered correctly. It is possible that learners could still perform better with

combination items if there were more eliciting stop items. As for word stress, although it led to significantly less successful uptake than eliciting stop (ASR value is -3.1 against 1.2), it resulted in 76.0% of correct answers, 9.3% higher than eliciting stop (rising intonation was excluded from the chi-square analysis due to low expected cell counts). The higher percentage of correct answers for word stress could be because learners had paid attention to teacher feedback when they simply acknowledged the feedback or responded to it inconclusively in episodes with word stress.

General extralinguistic cues refer to any body movement that accompanies verbal feedback. The word "general" is also used to mean that the term does not concern itself with any specific kind of bodily cues. Chi-square analysis indicates that there is a statistically significant relationship between extralinguistic cues and the occurrence of uptake. Compared with episodes without extralinguistic cues, episodes with extralinguistic cues led to slightly more uptake (65.9% versus 64.7%) and no uptake (23.1% versus 20.1%) but a significantly lower level of no opportunity (ASR=-2.3). To put it differently, learners had significantly more opportunities to respond to teachers in episodes with extralinguistic cues, and yet they only produced slightly more uptake. On more occasions they chose not to respond to teachers even when there was a chance (ASR=1.4). With that said, it can be concluded that learners in the present study tended to be less responsive to teacher feedback in episodes with extralinguisitc cues. This is totally different from Davies' (2006) finding that episodes with extralinguistic cues were likely to result in more uptake than topic continuation (no opportunity). One reason for this difference is that Davies classified learners' responses to teacher feedback in a different

way. In Davies' study, learner responses were coded as "either uptake or topic continuation, that is, when a learner notices an error and produces pushed output of an acceptable form or when a learner fails to notice and continues with the topic" (*p.* 843). His topic continuation refers to learners' continuation of a topic, which would most likely be considered as no uptake in the present study. Moreover, Davies did not distinguish between no uptake and no opportunity. It seemed that teachers' topic continuation which prevented learners from responding was not considered. In the current study, such episodes were coded as no opportunity. These coding and operational differences between the two studies could have influenced the calculation of different types of learner uptake, thus leading to different conclusions. A second important reason is that Davies included peer-repair as part of learner uptake while the current study only included the response by learners who were involved in an episode. The exclusion of peer repair in learner uptake could have significantly reduced the amount of learner uptake in the present study.

With respect to the successfulness of uptake, extralinguistic cues were shown to have a statistically significant effect on it (p=.005). Compared with episodes without extalinguistic cues, episodes with extralinguistic cues led to significant less successful uptake (ASR=-3.5) but significantly more acknowledge (ASR=2.9). It is difficult to tell why this is so. One possible explanation is that more than half of extralinguistic cues (51.9%) were used to explain word meanings in message-related episodes. Such episodes, as have been repeatedly discussed, led to more acknowledge. Another possible explanation is that teachers sometimes did not give full verbal feedback when

extralinguistic cues clearly conveyed a meaning. The feedback is therefore partially linguistic and partially extralinguistic. This could have made it more difficult for learners to incorporate the target feature in their language production. Learners hence may choose to simply acknowledge it or write it down. For example,

(88) Biey: What is disperse?
T: Disperse, ok. What do you think, so, look at the sentence there. Who dispersed? Police dispersed who?
Biey: The crowd.
T: The crowd. Ok, it could work either way. So what did the police want? They want the crowd together or apart? Disperse. (Hand gestures "together" and "apart" and then "disperse")
Biey: Oh ok.

In this example, the teacher hand gestures the meaning of "disperse" without really giving the learner a clear verbal definition of the word. The learner might have an idea that it somewhat carries the meaning of the word "apart", but it might be difficult for her to uptake the feedback by giving "disperse" a clear verbal definition herself or to use it in an utterance. As a result, she simply acknowledges it.

The lower level of uptake and successful uptake of episodes with extralinguistic cues might lead one to expect that these episodes would result in less noticing and learning. Analysis of the stimulated recall comments and test results, however, shows that the reverse is true. Adjusted standardized residuals indicate that learners talked significantly more about teacher feedback after viewing episodes with extralinguistic cues than after viewing episodes without extralinguistic cues (ASR value is 2.6 versus -2.6). The percentage of teacher-feedback-related comments of the former is 14.5% higher than that of the latter. In both the immediate test and delayed test, learners did significantly better with test items developed from episodes with extralinguistic cues than those developed

from episodes without extralinguistic cues (ASR value is 3.0 versus -3.0 in the immediate test and 2.3 versus -2.3 in the delayed test). The percentage of correct answers of the former is 17.3% higher than that of the latter in the immediate test and 13.5% higher in the delayed test. These results show that learners had paid more attention to teacher feedback when extralinguistic cues were used and as a result more learning had happened. Talking about the reasons why gestures, one type of extralinguistic cues, are relevant to the study of second language acquisition, Gullberg (2006) argues that gestures constitute input and thus can enhance learning. Hostetter and Alibali (2004, cited in Gullberg, 2006) observe that gestures help capture attention, provide redundancy, or engage more senses by grounding speech in the concrete, physical experience. These claims, although specifically about gestures, can be extended to extralinguistic cues in general now that other types of extralinguistic cues also constitute input and have similar properties. The observations by Gullberg and Hostetter and Alibali therefore can also explain the results of the recall interviews and the two tests in the present study.

Unlike general extralinguistic cues, the type of extralinguistic cues is concerned with different kinds of bodily nonlinguistic cues. Chi-square analysis indicates that the type of extralinguistic cues exerted a statistically significant effect on the occurrence of uptake. Adjusted standardized residuals reveal that among the five categories of extralinguistic cues, head movements led to significantly more uptake (ASR=3.7) while gestures led to significantly less uptake (ASR=-2.6). This seems to show that learners responded to teachers more actively in episodes with head movements than in episodes with gestures. A look at the category of no opportunity, however, shows that episodes with gestures

have the highest level of no opportunity while episodes with head movements have the lowest level of no opportunity (ASR value is 4.5 against -2.5). This result points to the possibility that the low level of uptake in episodes with gestures had been caused by the lack of opportunity.

Another notable thing is that there is a significantly higher level of no uptake in episodes with the combination of different types of extralinguistic cues (ASR=2.1) than episodes with any other single type of extralinguistic cues. It is not clear why this is so. One possibility is that learners understood the feedback so well with the help of multiple types of extralinguistic cues that they felt it unnecessary to respond to it. This may sound irrational. Learners' recall comments, however, show that it is possible. For example, after viewing an episode with multiple extralinguistic cues, a learner stated with a frown: "I don't think nothing because I was bored. She explained it too much." The learner did not clearly say what made the explanation "too much", but the multiple extralinguistic cues could be part of it.

In terms of the successfulness of uptake, it was not significantly affected by the type of extralinguistic cues. According to the percentages of successful uptake, head movements resulted in the highest level of successful uptake (51.4%) while gestures led to the lowest level of successful uptake (33.3%). The detailed transcription of the observation data shows that head movements are either nodding or headshaking. Nodding was mostly used when teachers were confirming a learner's correct production or language hypothesis while headshaking was mostly used when teachers wanted to emphasize a crucial point in

their feedback. In either case, learners' responses to the feedback were often successful. Below are two examples where teachers use head movements in their feedback.

(89) Lofa: Mrs. H.
T: Yes?
Lofa: Is my a pronoun?
T: Mm-hmm.
Lofa: Why a pronoun?
T: Mm, it is a pronoun. It's a possessive pronoun. Yeah.
Lofa: Possessive pronoun.
T: Yeah, yeah, yeah. (Nods)

(90) Ne: (To T) So neither just for the future.
T: No no no, if the sentence is affirmative, I am going to study English, you say me too. I am also going to study English.
Ne: (Nods)
T: If I say I am NOT (Stresses "not" and shakes head) going to New York city...
Ne: Me neither.
T: Me neither. I am not going to New York.

In the first episode, the teacher nods to confirm the learner's correct response to her feedback. In the second episode, the teacher shakes head to help stress a negation which is crucial to the learner's understanding of the target feature. Upon that, the learner provides the correct form before the teacher finishes her sentence. In both cases, learners' responses to teacher feedback were coded as successful uptake. Gestures, on the other hand, were used to explain word meanings 57.2% of the time. The feedback as discussed above might be more difficult for learners to uptake because gestures themselves partially took the place of verbal explanation and no direct definitions of the words were provided. This could explain why gestures resulted in a higher combined percentage of acknowledge and inconclusive uptake (66.7%) than head movements (48.6%). Given the situations where head movements and gestures were used, it is understandable that the former resulted in more successful uptake.

With respect to the stimulated recall interviews, no statistical analysis was performed due to low expected cell counts. What can be found from the available descriptive statistics is that learners talked more about teacher feedback after viewing episodes with head movements. Given that learners produced significantly more (successful) uptake and hence might be more responsive in episodes with head movements, this is hardly surprising. One noticeable thing is that episodes with the combination of different types of extralinguistic cues, which led to the highest level of no uptake and a low percentage of successful uptake next to gestures, have a percentage of teacher-feedback-related comments as high as that of episodes with head movements (80.0%). This means that learners actually were paying as much attention to teacher feedback in these episodes as in those with head movements. Another thing that stands out is that gestures, which have both the lowest level of uptake and successful uptake, also resulted in 76.4% of teacher-feedback-related comments, only a few percent lower than head movements and combination. This once again proves that uptake and successful uptake are not the only indicators for noticing.

With regard to test results, the type of extralinguistic cues did not show any statistically significant effect on either test. Descriptive statistics reveal that learners did best (73.6% of correct answers) with test items developed from episodes with gestures in the immediate test. The percentage of correct answers is close for items developed from episodes with the combination of different types of extralinguistic cues (71.1%). Items from episodes with head movements, on the other hand, have the lowest percentage of correct answers (63.7%). In the delayed test, gestures and combinations still have a

higher percentage of correct answers than head movements (54.3%, 75.8%, and 52.4% respectively). The higher percentages of correct answers for gestures and combinations correspond to the fact that both the two have a higher percentage of teacher-feedback-related comments, proving that more noticing leads to more learning.

What is surprising is that head movements led to the highest level of uptake, successful uptake, and teacher-feedback-related comments and yet resulted in the lowest percentage of correct answers in both tests. One possible explanation for this result is that the noticing aroused by head movements tends to only have a temporary effect. If so, it could be because head movements, both head shaking and nodding, are among the most brief and common extralinguistic cues and hence might not be as striking and impressive as other more complicated but less used extralinguistic cues.

Another thing that deserves special attention is the difference between gestures and combinations in terms of the change of correct answers in the two tests. While the correct answers for test items developed from episodes with gestures dropped by 19.3%, the correct answers for items developed from episodes with a combination of different types of extralinguistic cues increased by 4.7%. This could be because combinations with various extralinguistic cues aroused learners' attention to a deeper level and hence better helped them retain their memory while gestures as a single type of extralinguistic cues did not have an effect as long lasting as that of multiple extralinguistic cues. As for the reason why there is an increase in correct answers for items developed from episodes

with multiple extralinguistic cues, it could be because such episodes have a delayed effect on test results (for more discussion, see Gass, 1997; Mackey, 1999).

Type of gestures is a subcategory of extralinguistic cues which specifically refer to the movement of the hand and arm. It was found that although the type of gestures significantly affected the results of the immediate test (p=.029), there is no statistically significant difference among the different types of gestures in terms of the occurrence of uptake, the successfulness of uptake, or the delayed test (p>.05 in all three cases). About the stimulated recall comments, no chi-square analysis was conducted due to low expected cell counts.

According to the descriptive statistics, two types of gestures distinguish themselves from others: metaphorics and beats. Metaphorics led to the highest percentage of uptake (68.4%) and yet a lower (but not the lowest) percentage of successful uptake (34.6%). Metaphorics present "an image of the invisible -- an image of the abstraction" (McNeil, 1992, p.14). Even though they are pictorial, the contents are still abstract. Learners might find it more difficult to understand the abstract contents and therefore were pushed to respond to teachers, thus producing more uptake. On the other hand, exactly for the same reason, learners might not be so sure about their own understanding of the contents and therefore were reluctant to incorporate the target features in their language production, thus leading to less successful uptake.

In terms of the stimulated recall comments, it was found that learners paid more attention to teacher feedback when they felt they did not understand a language feature (e.g., "I was listening to the teacher's explanation because I didn't know what to call the stick either"; "I didn't know economical means from money perspective, so I paid special attention to it when the teacher explained it"). Now that metaphoric gestures were used to explain abstract ideas or concepts that are more difficult to understand, learners should pay more attention to teacher feedback with this type of cues. Results from the stimulated recall interviews, however, show that learners made the lowest percentage of comments on teacher feedback and the highest percentage of comments about other issues after viewing episodes with metaphoric gestures. It is difficult to tell if this is mere chance or for some unknown reasons due to the lack of inferential statistics. A possible explanation is that as a result of a less thorough understanding of the abstract contents, learners also found it difficult to talk about the feedback and therefore chose to talk about other issues concerning themselves (e.g., what they were guessing, how they felt about the new language feature, and whether they were checking a dictionary). The low percentage of comments on teacher feedback in relation to metaphoric gestures therefore may not mean learners did not pay much attention to the feedback.

Regarding the test results, learners provided the highest percentage of correct answers for test items developed from episodes with metaphoric gestures in both the immediate test and delayed test. If attention is a necessary condition for learning, this result proves that learners not only paid attention to teacher feedback with metaphoric gestures, they also paid more attention to it.

Another type of gestures that proved intriguing is beats. A glimpse at the results shows that the percentage of uptake for episodes with beats (66.7%) is very close to the percentage for episodes with metaphoric gestures. The percentage of successful uptake for episodes with beats (50.0%), however, is much higher than the percentage for the latter. With respect to the stimulated recall comments, episodes with beats led to 80% of teacher-feedback-related comments, 17.5% higher than episodes with metaphoric gestures. These results as a whole mean that learners might have paid more attention to teacher feedback in episodes with beats. Learners' performance with items developed from such episodes therefore should also be better in the two tests. In the delayed test, beats were excluded from the chi-square analysis due to low expected cell counts. Results from the immediate test show that learners actually provided the lowest percentage of correct answers for items developed from episodes with beats (40.0%), 50% lower than those developed from episodes with metaphoric gestures. One possible explanation for this surprising result comes from the nature of beats. McNeil (1992) notes that "Of all gestures, beats are the most insignificant looking" (p. 15) despite its semiotic value. This "insignificant look" might have impaired the ability of beats to help learners retain their memory even though they might have helped learners notice teacher feedback at the moment.

5.8 Summary

In summary, the results reported in this chapter suggest that the occurrence of uptake was significantly affected by general paralinguistic cues, general extralinguistic cues, and the

type of extralinguistic cues. The presence of general paralinguistic cues, the presence of general extralinguistic cues (to a lesser degree), and the use of head movements led to more uptake. The successfulness of uptake was significantly affected by general paralinguistic cues, the type of paralinguistic cues, and general extralinguistic cues. Learners produced a higher level of successful uptake when general paralinguistic cues were present, when rising intonation was used, and when general extralinguistic cues were absent. Learners' stimulated recall comments were found to be significantly affected by general extralinguistic cues. The presence of general extralinguistic cues resulted in a higher rate of teacher-feedback-related comments. It is not clear whether other variables have an effect on the comments due to lack of inferential information. About testing, the results of the immediate test were significantly affected by general extralinguistic cues and the type of gestures. The presence of general extralinguistic cues and the use of metaphoric gestures predicted more correct answers. The results of the delayed test were only significantly affected by general extralinguistic cues. Similar to the immediate test, learners performed better with test items developed from episodes with extralinguistic cues in teacher feedback. These results taken as whole show that the answer to the second research question is: yes, the nonlinguistic cues in teacher feedback affect learners' noticing and learning to a certain degree, especially at the general level.

CHAPTER 6

METALANGUAGE, THE NOTICING AND THE EFFECT OF TEACHER FEEDBACK

6.1 Introduction

This chapter addresses the third research question: Does the metalanguage in teacher feedback affect learners' noticing and learning? After presenting the frequency of general metalanguage and the distribution of different types of metalanguage, I report and discuss the relationship between metalanguage and learners' responses to teachers' feedback, learners' comments in the stimulated recall interviews, and results from both the immediate test and delayed test. By metalanguage I mean the use of technical and non-technical linguistic terms such as "adjective" and "spelling".

6.2 Occurrence and distribution of metalanguage

This section briefly reports the frequency of general metalanguage and the distribution of different types of metalanguage. Table 75 shows that some sort of metalanguage, be it technical or non-technical, was used in 71.3% of a total of 1434 feedback episodes. Table 76 shows that among a total of 1023 feedback episodes where metalanguage was used, 7.2% only carry technical terms; 60.2% only contain non-technical terms; and 32.6% have both technical and non-technical terms.

Table 75 Occurrence of general metalanguage

	Number	Percentage
Present	1023	71.3%
Absent	411	28.7%
Total	1434	100%

Table 76 Distribution of different types of metalanguage

	Number	Percentage
Tech-only	74	7.2%
Non-tech-only	616	60.2%
Tech+non-tech	333	32.6%
Total	1023	100%

Table 77 shows the distribution of 53 randomly selected terms according to their properties and the number of teachers who used them in their feedback to students. A quick look at the table reveals that terms that were used by more teachers are often non-technical while terms that were used by fewer teachers are often technical. For example, among the terms that were used by all eight teachers, only one is technical (noun); and among the terms that were used only by one teacher, all are technical.

Table 77 Randomly selected teacher metalanguage by property and number of teachers

Terms	Properties	Examples	N of teachers
Call	Non-tech	T: When you do it very loudly, that's **called** "slam", ok?	8
Mean/meaning	Non-tech	T: No, scram **means** to leave fast.	8
Noun	Tech	T: So every adjective needs a **noun**, so American what?	8
Say	Non-tech	T: Why does water boil? That's the way we want to **say** it.	8
Spell/spelling	Non-tech	T: And the way it's **spelled** now, it's already lumped together in one word.	8
Use	Non-tech	T: You CAN **use** *there* in a tag question.	8
Need	Non-tech	T: I don't think you **need** to use the.	7
Same	Non-tech	T: Yeah, they kind of have the **same** meaning, but the way that you used them are a little bit different.	7
Talk about	Non-tech	T: Sometimes we use this one to **talk about** religion.	7
Verb	Tech	T: Thank you for saying that 'cause treat can be a **verb**.	7
Adjective	Tech	T: Adjective, necessary is an **adjective**.	7
Describe	Non-tech	T: It's fog right? You can **describe** today as a foggy day.	6
Sound	Non-tech	T: It **sounds** the same, but the very you're thinking of has an E...but this one here is a verb, meaning things can be different.	6
Subject	Tech	T: Ok, your **subject** is it. So this meaning of look is not an action verb.	6

Table 77 (cont'd)

Way	Non-tech	T: The **way** it's going to be used is, they're going to say by Jove.	6
Idea	Non-tech	T: Ok, the **idea** of hold back is to eh not to go forward, all right?	6
Make	Non-tech	T: T: How about this one? Can you **make** it into a question?	6
Part	Non-tech	T: So, a sentence with two **parts**, we have clauses.	6
Singular	Tech	T: InhabitantS because people are plural or **singular**?	5
Tense	Tech	T: Like add –ed or…Change the **tense**.	5
Adverb	Tech	T: Ok, again we have an **adverb**. Don't forget the **adverb**.	5
Go+preposition	Non-tech	T: So it should **go between** pick and up because that verb is a phrase.	5
Name	Non-tech	T: T: To plus verb. Do you know the **name**?	5
Present	Non-tech	T: But this one we can either use to talk about the past…or to talk about the **present** because we can say I am used to studying at night.	4
Question	Non-tech	T: When you ask a **question** about the object, what do you need?	4
Writing	Non-tech	T: My example is for speaking, but the grammar is for **writing** also.	4
Answer	Non-tech	T: Possibly we usually use like a short **answer**.	4
Letter	Non-tech	T: No, S-U-E, three **letters**. Yeah.	4
Object	Tech	T: T: We, we don't say cut corners and then an **object**.	4
Phrase	Non-tech	T: So do you know this **phrase** put down roots?	4
Syllable	Tech	T: HealthiER right? It's two **syllables** and ends in y.	3
Term	Non-tech	T: No, it's a legal **term**.	3
Base form	Tech	T: What's before the **base form**?	3
Intonation	Tech	T: But there is a question, so you got to do the question **intonation**.	3
Past tense	Tech	T: We need **past tense**, **past tense**. You said she gets.	3
Vowel	Tech	T: Brevity. Totally different, different **vowel** sound,	3
Compound noun	Tech	T: No, househusband is just a **compound noun**.	2
Count noun	Tech	T: It's not a **count noun**, so you can't say garbages. Just garbage.	2
Helping verb	Tech	T: You're missing your **helping verb** in the first one.	2
Neutral	Non-tech	T: This is more common, more **natural**.	2
Noun phrase	Tech	T: So sometimes it's a whole **noun phrase**, not just a noun.	2
Part of speech	Tech	T: Recently. Ok, what **part of speech** is recently?	2
Prefix	Tech	T: What does pre- mean? Pre- is a **prefix**.	2
Pronoun	Tech	T: Ok, where is the **pronoun** of the sentence?	2
Simple present	Tech	T: No no no. No, all people die. **Simple present**.	2
Stress	Tech	T: Mm-hmm. Same **stress**, economical.	2

Table 77 (cont'd)

Suffix	Tech	T: Informality or formality. We put that stress right before the **suffix**.	1
Synonym	Tech	T: Yeah, kind of like complain. That's a good **synonym**.	1
Past participle	Tech	T: That's a **past participle**.	1
Past progressive	Tech	T: But let's just try **past progressive** and simple past.	1
Simple past	Tech	T: I don't know, but I like the way you said that, with the **simple past**.	1
Tag (question)	Tech	T: Sometimes, it doesn't have to have a negative in a **tag question**.	1
Uncount	Tech	T: Less homework, **uncount**.	1

6.3 Metalanguage and learner uptake

This section reports the relationship between the metalanguage in teacher feedback and learner uptake. The occurrence and successfulness of uptake are presented together according to both general metalanguage and different types of metalanguage.

General metalanguage

Tables 78 and 79 illustrate the occurrence and successfulness of uptake according to the presence of general metalanguage in teachers' feedback. Table 78 shows that learners responded to teacher feedback in 65.5% of episodes where teachers used metalanguage and 64.5% of episodes where teachers did not use metalanguage. Adjusted standardized residuals in the table reveal that while there is no significant difference between the two in terms of uptake and no uptake, no opportunity was significantly more frequent in episodes where teachers did not use metalanguage. Table 79 shows that when uptake did occur, 58.1% was successful in episodes where metalanguage was absent from teacher feedback and 40.1% was successful in episodes where metalanguage was present in teacher feedback. Adjusted standardized residuals in the table reveal that episodes with

metalanguage in teacher feedback led to significantly less successful uptake and significantly more inconclusive feedback. Chi-square tests indicate that both the occurrence and successfulness of uptake are significantly related to the presence of general metalanguage (*p* value is .021 for the occurrence of uptake and .000 for the successfulness of uptake).

Table 78 General metalanguage and the occurrence of uptake

	Number of episodes	Occurrence of uptake		
		Uptake	No uptake	No opportunity
Present	123	670 (65.5%)	230 (22.5%)	123 (12.0%)
Residuals (Present)	--	.4	1.7	-2.5
Absent	70	265 (64.5%)	76 (18.5%)	70 (17.0%)

$X^2(2, n=1434)=7.700, p=.021$

Table 79 General metalanguage and the successfulness of uptake

	Number of uptake	Successfulness of uptake			
		Successful	Unsuccessful	Acknowledge	Inconclusive
Present	670	269 (40.1%)	13 (1.9%)	266 (39.7%)	122 (18.2%)
Residuals (Present)	--	-4.8	.4	1.9	3.9
Absent	265	154 (58.1%)	4 (1.5%)	86 (32.5%)	21 (7.9%)

$X^2(3, n=935)=28.368, p=.000$

Type of metalanguage

Tables 80, 81, and 82 illustrate the occurrence and successfulness of uptake according to the type of metalanguage. Table 80 shows that learners responded to teacher feedback in 75.7% of episodes with technical terms only, 65.9% of episodes with non-technical terms only, and 62.5% of episodes with technical plus non-technical terms. Adjusted standardized residuals in Table 81 reveal that while uptake was more frequent (although not significantly) in episodes where teachers used technical terms only, no uptake was

significantly more frequent in episodes where teachers used technical plus non-technical terms and no opportunity was significantly more frequent in episodes where teachers used non-technical terms only. Table 82 shows that when uptake did occur, 44.6% was successful in episodes where teachers used technical terms only, 40.6% was successful in episodes where teachers used non-technical terms only, and 38.0% was successful in episodes where teachers used both technical and non-technical terms. Chi-square tests indicate that while the occurrence of uptake is significantly related to the type of metalanguage, the successfulness of uptake is not (p value is .027 for the occurrence of uptake and .319 for the successfulness of uptake).

Table 80 Type of metalanguage and the occurrence of uptake

	Number of episodes	Occurrence of uptake		
		Uptake	No uptake	No opportunity
Tech-only	74	56 (75.7%)	13 (17.6%)	5 (6.8%)
Non-tech-only	616	406 (65.9%)	126 (20.5%)	84 (13.6%)
Tech+non-tech	333	208 (62.5%)	91 (27.3%)	34 (10.2%)

$X^2(2, n=1023)=11.004, p=.027$

Table 81 Type of metalanguage and the occurrence of uptake residuals

	Uptake	No uptake	No opportunity
Tech-only	1.9	-1.1	-1.4
Non-tech-only	.3	-1.9	2.0
Tech+non-tech	-1.4	2.6	-1.2

Table 82 Type of metalanguage and the successfulness of uptake

	Number of uptake	Successfulness of uptake			
		Successful	Unsuccessful	Acknowledge	Inconclusive
Tech-only	56	25 (44.6%)	0 (0.0%)	17 (30.4%)	14 (25.0%)
Non-tech-only	406	165 (40.6%)	6 (1.4%)	164 (40.4%)	71 (17.5%)
Tech+non-tech	208	79 (38.0%)	7 (3.4%)	85 (40.9%)	37 (17.8%)

$X^2(6, n=670)=7.020, p=.319$

To sum up, the descriptive and inferential analysis in this section reveals that general metalanguage exerted a statistically significant effect on both the occurrence and successfulness of uptake ($p<.05$). The presence of metalanguage in teacher feedback led to slightly more learner uptake (ASR=.4) but significant less successful uptake (ASR=-4.8). The type of metalanguage also significantly affected the occurrence of uptake but did not have any statistically significant effect on the successfulness of uptake. Adjusted standardized residuals reveal that when the metalalanguage in teacher feedback was all technical, learners produced more uptake. The value of the adjusted standardized residual is not up to 2.0 but very close (ASR=1.9).

6.4 Metalanguage and recall comments

This section reports the relationship between the metalanguage in teacher feedback and learners' comments in the stimulated recall interviews. The frequency of comments related to teacher feedback is presented in relation to both general metalanguage and different types of metalanguage.

General metalanguage

Table 83 shows the rate of teacher-feedback-related comments in relation to the presence of general metalanguage in teacher feedback. A total of 72.2% of recall comments are related to teacher feedback for episodes where teachers used metalanguage and 70.6% are related to teacher feedback for episodes where teachers did not use metalanguage. Chi-square test results suggest that the rate of teacher-feedback-related comments was not

significantly affected by the presence of general metalanguage in teacher feedback (p=.700).

Table 83 General metalanguage and teacher-feedback-related comments

	Total number of comments	Teacher-feedback-related	Non-teacher-feedback-related	Other
Present	245	177 (72.2%)	25 (10.2%)	43 (17.5%)
Absent	34	24 (70.6%)	5 (14.7%)	5 (14.7%)

X^2(2, n=279)=.715, p=.700

Type of metalanguage

Table 84 shows the rate of teacher-feedback-related comments in relation to the type of metalanguage. A total of 57.1% of recall comments are related to teacher feedback for episodes where teachers only used technical terms, 74.8% are related to teacher feedback for episodes where teachers only used non-technical terms, and 69.4% are related to teacher feedback for episodes were teachers used both technical and non-technical terms. Chi-square test results suggest that the rate of teacher-feedback-related comments was not significantly affected by the type of metalanguage (p=.303).

Table 84 Type of metalanguage and teacher-feedback-related comments

	Total number of comments	Teacher-feedback-related	Non-teacher-feedback-related	Other
Tech-only	14	8 (57.1%)	1 (7.1%)	5 (35.7%)
Non-tech-only	159	119 (74.8%)	14 (8.8%)	26 (16.4%)
Tech+non-tech	72	50 (69.4%)	10 (13.9%)	12 (16.7%)

X^2(4, n=245)=4.854, p=.303

To sum up, the descriptive and inferential analysis in this section reveals that neither general metalanguage nor the type of metalanguage exerted a statistically significant

effect on the rate of teacher-feedback-related comments (p value is .700 for general metalanguage and .303 for the type of metalanguage).

6.5 Metalanguage and test results

This section reports the relationship between the metalanguage in teacher feedback and test results. Results from both the immediate test and delayed test are presented together according to general metalanguage and different types of metalanguage.

General metalanguage

Table 85 is an illustration of test results according to the presence of general metalanguage in teacher feedback. In the immediate test, the percentage of correct answers is 66.8% for episodes with metalanguage in teacher feedback and 56.4% for episodes without metalanguage in teacher feedback. In the delayed test, the percentage of the two is 54.8% and 47.7% respectively. Chi-square analysis indicates that the presence of general metalanguage did not significantly affect the results of either test (p value is .147 for the immediate test and .311 for the delayed test).

Table 85 General metalanguage and test results

	Immediate test			Delayed test		
	Number of test items	Correct	Incorrect	Number of test items	Correct	Incorrect
Present	217	145 (66.8%)	72 (33.1%)	217	119 (54.8%)	98 (45.1%)
Absent	55	31 (56.4%)	24 (43.6%)	65	31 (47.7%)	34 (52.4%)
	$X^2(1, n=272)=2.101, p=.147$			$X^2(1, n=282)=1.026, p=.311$		

Type of metalanguage

Table 86 is an illustration of test results according to the type of metalanguage in teacher feedback. In the immediate test, the percentage of correct answers is about 67.0% for episodes with all the three types of metalanguage. In the delayed test, the percentage of correct answers is 66.7% for episodes with technical terms only, 44.9% for episodes with non-technical terms only, and 69.2 for episodes with both technical and non-technical terms. Adjusted standardized residuals suggest that learners provided significantly more correct answers for test items developed from episodes with both technical and non-technical terms and significantly less correct answers for items developed from non-technical terms only. Chi-square analysis indicates that the type of metalanguage significantly affected the results of the delayed test (p=.002) but not the results of the immediate test (p=.998).

Table 86 Type of metalanguage and test results

	Immediate test			Delayed test			
	Number of test items	Correct	Incorrect	Number of test items	Correct	Residuals (Correct)	Incorrect
Tech-only	9	6 (66.7%)	3 (33.3%)	12	8 (66.7%)	.8	4 (33.3%)
Non-tech-only	135	90 (66.7%)	45 (33.3%)	127	57 (44.9%)	-3.5	70 (55.1%)
Tech+non-tech	73	49 (67.1%)	24 (32.8%)	78	54 (69.2%)	3.2	24 (30.8%)
	$X^2(2, n=217)=.005, p=.998$			$X^2(2, n=217)=12.285, p=.002$			

To sum up, the descriptive and inferential analysis in this section reveals that general metalanguage did not have any statistically significant effect on the results of either test (p>.05). Concerning the type of metalanguage, while it did not significantly affect the results of the immediate test, it did show a statistically significant effect on the results of

the delayed test (p=.002), with episodes where both technical and non-technical terms were used leading to significantly more correct answers (ASR=3.2).

6.6 Review of results

The results presented above show that the occurrence of uptake was significantly affected by both the presence of general metalanguage and the type of metalanguage but the successfulness of uptake was only significantly affected by the former, not the latter. In terms of the rate of teacher-feedback-related recall comments, neither the presence of general metalanguage nor the type of metalanguage had a statistically significant effect on it. This is also the case with the results of the immediate test. The results of the delayed test, on the other hand, were significantly affected by the type of metalanguage but not by the presence of general metalanguage. Despite the fact that both general metalanguage and the type of metalanguage had a significant effect on the occurrence of uptake and the former also significantly affected the successfulness of uptake, no category predicted more uptake or more successful uptake. The correct answers in the delayed test, however, were predicted by episodes with technical plus non-technical terms. The table below is a summary of the overall results in this section.

Table 87 Overall results by metalanguage

	Occurrence of uptake	Successfulness of uptake	Feedback-related comments	Immediate test	Delayed test
General metalanguage	+	+	-	-	-
Type of metalanguage	+	-	-	-	+

6.7 Discussion:

This section discusses the results presented in the previous sections. The occurrence and successfulness of uptake are examined together with learners' stimulated recall comments and test results in relation to general metalanguage and the type of metalanguage.

As reviewed in Chapter 2, very little empirical research has been conducted to specifically examine the effect of teachers' metalanguage in second language classrooms. Borg's (1998, 1999) studies were mainly concerned with teachers' beliefs about metalanguage use. They therefore did not speak directly to the effect of metalanguage. One group of studies which did touch upon metalanguage (e.g., Lyster & Ranta, 1997; Panova & Lyster, 2002; Sheen, 2004) is those which investigated the effect of different types of feedback. A general finding of these studies is that metalinguistic feedback, a feedback type which contains metalanguage, is more successful than recasts in eliciting learner uptake and repair. Results from the present study provide both evidence and counterevidence for this finding. It was found that the presence of metalanguage in teacher feedback did lead to more learner uptake but only slightly. On the other hand, there was significantly less successful uptake when teachers used metalanguage in their feedback. This difference could have resulted from different teaching contexts, different participants, and different research designs. Meanwhile, the lower level of uptake and successful uptake in the present study somewhat echoes the findings by Basturkmen et al. (2002). In that study, the researchers examined the effectiveness of metalanguage by different types of focus-on-form episodes. The results they reported show that uptake and successful uptake co-occurred with metalanguage mostly frequently in student-initiated

focus-on-form episodes, but the percentage was not very high (50.3% for uptake and 44.2% for successful uptake). Unlike the present study, Basturkmen et al. did not investigate the relationship between the specific types of metalanguage and learner uptake. This is discussed below in addition to general metalanguage.

General metalanguage refers to any language that is used to describe the language system. The word "general" again is used to mean that the term does not concern itself with any specific kind of metalinguistic terms. Chi-square analysis shows that the presence of metalanguage had a significant effect on both the occurrence and successfulness of uptake (*p* value is .021 and .000 respectively). In terms of the occurrence of uptake, the percentage of uptake is virtually the same whether or not there was metalanguage in teacher feedback (65.5% and 64.5% respectively). On the other hand, learners had significantly more opportunities to respond to teacher feedback when metalanguage was present (the adjusted standardized residual of no opportunity is -2.5) and yet these episodes resulted in a higher level of no uptake although it is not statistically significant (ASR=1.7). Taken together, these results indicate that learners were less responsive to teacher feedback with the presence of metalanguage. This could be because some metalanguage (e.g., grammar labels) had exerted a threatening effect on learners as the teachers in Borg's (1998, 1999) studies were afraid, and had prevented them from responding to teacher feedback. It could also be because some metalanguage that learners were not familiar with had added to their cognitive burdens as some researchers have argued (e.g., Corder, 1973), and had also prevented learners from responding to the feedback.

In terms of the successfulness of uptake, adjusted standardized residuals indicate that learners produced significantly less successful uptake when metalanguage was used in teacher feedback (ASR=-4.8). However, it led to significantly more inconclusive uptake (ASR=3.9) and a much higher level of acknowledge (ASR=1.9). This could also be because metalanguage was threatening to learners and/or had added to their cognitive burdens so that they decided to respond to the feedback by simply acknowledging it or writing it down instead of verbally responding to it.

With respect to the stimulated recall comments, chi-square analysis indicates that the presence of metalanguage did not significantly affect the rate of teacher-feedback-related comments (p=.700). On the other hand, the descriptive statistics reveal that the percentage of teacher-feedback-related comments is slightly higher for episodes with metalanguage in teacher feedback than those without metalanguage in teacher feedback (72.2% versus 70.6% respectively). This means that learners could have paid more attention to feedback with metalanguage although the difference is not significant. It could be because some metalanguage which is substantially different from everyday language had helped to arouse learners' attention to the feedback with its markedness. Another possibility is that some of the language features in episodes with metalanguage were the focus of a lesson or unit (e.g., the grammar rules in grammar practice activities). They therefore received more attention from learners.

Regarding the test results, chi-square analysis indicates that the presence of metalanguage did not significantly affect the results of either the immediate test or the delayed test. The percentages of correct answers, however, reveal that learners performed better with test items developed from episodes with metalanguage in teachers' feedback than those developed from episodes with no metalanguage in teachers' feedback (66.8% versus 56.4% in the immediate test and 54.8% versus 47.7% in the delayed test). Given the role of attention in learning, this could be because learners had indeed paid more attention to teacher feedback with metalanguage as shown in teacher-feedback-related comments.

Unlike general metalanguage, the type of metalanguage is concerned with different kinds of metalinguistic terms. Chi-square analysis indicates that there is a statistically significant relationship between the type of metalanguage and the occurrence of uptake (p=.027). Specifically, episodes with only technical terms led to the highest level of uptake as compared with episodes with non-technical terms only and episodes with both technical and non-technical terms (75.7% versus 65.9% and 62.5%). This is somewhat surprising given that technical terms should be more threatening to learners than other types of metalanguage. A careful look at episodes with only technical terms shed some light on this puzzle. Metalanguage in a majority of these episodes (78.4%) is terms that occurred frequently across the nine classes (see Table 77), sometimes from both teachers and learners, and hence might be familiar to learners. These include terms about part of speech (e.g., noun, verb, adjective, adverb), number (e.g., plural, singular), simple verb tense (e.g., present tense, past tense), etc. Below are a few examples:

(91) Xin: (Writes "garage sell")
T: (Looks at cross puzzle) Sale, not sell. Sell is the verb but sale is the, the noun. Sale, S-A-L-E.
Xin: S-A-L-E. (Writes on paper)

(92) Ye: I were...
T: I WAS, were is plural. (Stresses "was")
Ye: Yeah, I was doing my homework.

(93) (Reading a story that happened in the past)
Cadi: You're smiling.
T: Let's keep it in past tense.
Cadi: (Pauses, 4 seconds) You were smiling. (Makes changes on book)

In the first episode, the technical terms the teacher uses are "verb" and "noun", which can be found in the recordings of eight out of the nine classes. In the second episode, the teacher uses the term "plural", which the learner seems to understand by correcting herself in her new utterance after the feedback. In the third episode, the teacher uses the term "past tense", which the learner also seems to understand by incorporating the correct form in his new utterance. Some of the terms in these episodes have actually been used by learners themselves (e.g., "So collapsing is still a verb?" "T, garage sale, plural or single?"). Now that most of the technical terms are simple ones familiar to learners, learners might feel less threatened or burdened to respond to the feedback. In contrast to episodes with technical terms only, many episodes with both technical and non-technical terms cover complicated and infrequent terms too (e.g., "You CAN use 'there' in a tag question," "...but if you look at your list of non-action verbs, the word 'think' is not there...but in the extended present, you know...). These terms could have had an intimidating effect on learners who may have been reluctant to respond to the teachers' feedback.

About the successfulness of uptake, Chi-square analysis indicates that it was not significantly affect by the type of metalanguage (p=.319). Nonetheless, the descriptive statistics show that episodes with only technical terms resulted in a higher percentage of successful uptake (44.6%) than those with both technical and non-technical terms (38.0%). This difference could also be explained by the above-mentioned nature of the metalanguage in the two types of episodes. It is natural that learners would be more likely to be able to repeat the feedback or incorporate it in their own production when they were more familiar and felt more comfortable with the language in the feedback.

Although learners produced more (successful) uptake in episodes with technical terms only, they did not seem to have paid more attention to teacher feedback in these episodes. Analysis of the stimulated recall comments indicates that episodes with technical terms only actually resulted in a lower percentage of teacher-feedback-related comments than those with both technical and non-technical terms (57.1% versus 69.4%). This could be because the less familiar or unfamiliar terms in episodes with both technical and non-technical terms had helped arouse learners' attention since learners tended to pay more attention to teacher feedback when the target feature was new to them (see the discussion of Chapter 5). Another explanation is that teacher feedback in the majority of episodes with technical terms only (77.0%) is simpler, generally with one or two teacher turns. Attention to the target feature of these episodes was therefore brief.

The feedback in episodes with both technical and non-technical terms, by contrast, is more complicated. Among a total of 333 episodes, 63.7% have more than two teacher

turns. Attention to the target feature of these episodes therefore could be longer and maybe more intense. This also provides a good explanation for the reason why episodes with both technical and non-technical terms resulted in a higher percentage of correct answers than episodes with technical terms only in both the immediate test and delayed test. In the immediate test, the difference between the two is very slight (67.1% versus 66.7%) and not statistically significant (p=.998). In the delayed test, the percentage difference is still not big (69.2% versus 66.7%). However, adjusted standardized residuals show that episodes with both technical and non-technical terms resulted in significantly more correct answers than episodes with technical terms only (ASR value is 3.2 versus .8).

Up to now, the discussion about type of metalanguage has focused on the categories of technical terms only and technical plus non-technical terms. The other category, non-technical terms only, also provides some interesting results. In theory, feedback with non-technical terms only should be the least confusing, the least threatening, and the least burdening among the three types of metalanguage. Learners therefore should feel most comfortable to respond to teachers in episodes with such terms. In reality, however, episodes with only non-technical terms took the middle position in terms of both the occurrence and successfulness of uptake, with less uptake and successful uptake than episodes with technical terms only and more uptake and successful uptake than episodes with both technical and non-technical terms. To find out why there is a lower level of uptake in episodes with non-technical terms only, it is necessary to look at the category of no uptake and no opportunity. Adjusted standardized residuals reveal that among the

three types of metalanguage, non-technical terms only has a significantly higher level of no opportunity (ASR value is 2.0 against -1.4 for technical terms only and -1.2 for technical plus non-technical terms). On the other hand, it has a lower level of no uptake (ASR value is -1.9 against -1.1 for technical terms only and 2.6 for technical plus non-technical terms). Although the absolute value of its adjusted standardized residual is not up to the conventional 2.0, it is very close.

All this means that the lower level of uptake for non-technical terms only could have been the result of fewer opportunities, not because learners chose not to respond to teachers. There could be a variety of reasons why teachers gave learners fewer opportunities when the metalanguage in the feedback was non-technical terms only. One possibility is that teachers assumed that learners should have fewer problems with feedback without technical terms and therefore did not want to spend as much time on them. As for the reason why there is a lower level of successful uptake in episodes with non-technical terms only, it could be that learners, believing that they understood the feedback, thought they could afford not to use the target feature right after the feedback. They therefore simply acknowledged it or responded to it in an inconclusive way. Again, this could only be one of many possible explanations.

With regard to the stimulated recall comments, episodes with only non-technical terms in teachers' feedback resulted in the highest level of teacher-feedback-related comments (74.8% against 57.1% for technical terms only and 69.4% for technical plus non-technical terms). This means that learners had paid more attention to the feedback in

these episodes. Given that learners did not produce the highest level of uptake or successful uptake in episodes with non-technical terms only and that they had frequently responded to teacher feedback by acknowledging it or producing inconclusive uptake in these episodes, this result again points to the possibility that other types of uptake (e.g., acknowledge, inconclusive uptake) can also be indicators of attention. It should be noted that episodes with only non-technical terms in teacher feedback also resulted in some unsuccessful uptake, but the percentage is very low (1.4%). The very fact that learners tried to uptake the feedback is evidence that they were attending to it even though they failed to understand or use it.

Concerning the results of the two tests, they also look intriguing when associated with non-technical terms only. In the immediate test, learners' performance with test items developed from episodes with only non-technical terms in the feedback is almost the same with those developed from episodes with the other two categories of metalanguage. In the delayed test, however, learners provided significantly fewer correct answers for items developed from episodes with only non-technical terms (ASR value is -3.5 against .8 for technical terms only and 3.2 for technical plus non-technical terms). The percentage of correct answers for these items is more than 20% lower than its counterparts. Given that episodes with only non-technical terms resulted in the highest percentage of teacher-feedback-related comments, this is surprising. It is hard to pinpoint the exact reasons for this result, but a look at the nature of the attention learners paid to teacher feedback might help. When learners only acknowledged teachers' feedback or responded to it inconclusively, no pushed output was produced even though they might

have attended to the feedback. The attention learners paid to the feedback therefore might not be intense. As a result, this kind of uptake might not be as beneficial to learning as successful uptake, where learners tried to use the target feature in their language production. Consequently, learners' memory of the target features in episodes with acknowledge or inconclusive uptake decayed faster. If this were true, it is understandable that episodes with non-technical terms only, which resulted in more acknowledge and inconclusive uptake, would lead to less correct answers in the delayed test.

6.8 Summary

In summary, the results reported in this chapter suggest that the occurrence of uptake was significantly affected by two different factors: general metalanguage and the type of metalanguage. Episodes with general metalanguage and episodes with only technical terms resulted in a higher level of uptake. The successfulness of uptake was significantly affected by general metalanguage but not by the specific type of metalanguage. Learners produced more successful uptake when metalanguage was not present in teachers' feedback. Learners' stimulated recall comments did not have any significant relationship with either general metalanguage or the specific type of metalanguage. This is also the case with the results of the immediate test. The results of the delayed test, however, were significantly related to the type of metalanguage. Learners performed better with test items developed from episodes with both technical and non-technical terms. These results together show that the answer to the third research question is: yes, metalanguage in teacher feedback has some effect on noticing and learning.

CHAPTER 7

CONCLUSION

7.1 Introduction

This final chapter begins with a brief summary of the findings in previous chapters and a discussion of the implications of these findings. It continues with a reflection on the design of the study, and finally makes some tentative suggestions on future research.

7.2 Summary and implications

In this section, I briefly summarize the general findings of the present study according to the three research questions and discuss both the pedagogical implications and research implications of these findings.

7.2.1 Summary

Three questions are raised in this study:

1. Do the characteristics of teacher feedback episodes affect learners' noticing and learning?

2. Do the nonlinguistic cues in teacher feedback affect learners' noticing and learning?

3. Does the metalanguage in teacher feedback affect learners' noticing and learning?

In terms of the characteristics of feedback episodes, all seven characteristics examined affected all or some of the dependent variables that measured noticing and learning (the occurrence of uptake, the successfulness of uptake, learners' stimulated recall comments,

and test results). This also holds true for the presence of general nonlinguistic cues and some of their subcategories. As for teachers' metalanguage and its subtypes, they were found to have some effect on noticing and learning too although to a lesser degree and not always in a positive way.

7.2.2 Implications

The findings presented above carry important implications for both classroom teaching and feedback research. Although they may not apply to every teaching or research context, these implications can be very helpful for teaching and research in a context similar to that of the present study.

Implications for Teaching

As discussed in Chapters 4, 5, and 6, the findings from this study are not exactly the same as the findings in previous studies. It is therefore dangerous to make generalizations. However, some aspects of teacher feedback episodes have been repeatedly shown to be more effective than others in eliciting (successful) uptake, promoting noticing, and facilitating learning. Here I describe three findings that are reasonably plausible.

The characteristics of feedback episodes matter

The findings of this study align with previous studies to suggest that the characteristics of feedback episodes play a role in the noticing and effectiveness of teacher feedback. One example is that elicit episodes are more effective than provide episodes in drawing learner uptake and successful uptake in both immersion programs (Lyster, 1998a; Lyster

and Ranta, 1997) and ESL classrooms (Ellis et al. 2001a; Loewen, 2004; and also the present study). This means that learners might benefit more from the feedback if teachers try to prompt learners to get the correct language form instead of providing direct information to them. Another example is that complex episodes are more facilitative than simple episodes in increasing learners' noticing of teacher feedback and the subsequent learning. The findings in this study show that compared with simple episodes complex episodes resulted in significantly more uptake, successful uptake, teacher-feedback-related comments, and correct answers in both the immediate test and delayed test. This indicates that complex episodes are more effective in both drawing learners' attention to teacher feedback and their language improvement. Concerning characteristics whose effect is more arguable, teachers can make decisions according to the specific teaching context such as the teaching objective and student population.

Non-linguistic cues help

As for the use of nonlinguistic cues, there has been positive evidence from different research areas, for example, applied linguistics, speech communication, and second language acquisition too. Although nonlinguistic cues do not always help communication (See Faraco & Kida, 2008), they in general are effective in increasing learners' noticing of feedback and subsequent language learning. In the present study, the presence of general paralinguistic cues elicited significantly more uptake and successful uptake and the present of extralinguistic cues predicted significantly more teacher-feedback-related comments in the stimulated recall interviews and significantly more correct answers in both the immediate test and delayed test. These findings suggest that nonlinguistic cues

as a whole can promote learners' noticing of teacher feedback and the effectiveness of the feedback. Teachers therefore can incorporate nonlinguistic cues in their feedback to help improve students' learning.

Metalanguage is not always effective

Finally, about metalanguage use, the results of the present study as well as the findings by Basturkmen et al.' (2002) show that it can affect learners' noticing of teacher feedback and their language improvement in a certain way, but the overall effect is not very impressive or even counteractive. In the present study, the adjusted standardized residuals indicate that neither general metalanguage nor its subtypes predicted significantly more uptake, successful uptake, teacher-feedback-related comments, or correct answers in the immediate test. Only episodes with both technical and non-technical terms predicted significantly more correct answers in the delayed test. Moreover, when teachers used metalanguage in their feedback to learners, there was significantly less successful uptake, indicating that learners were significantly less able to incorporate the correct language forms in their new language production after receiving teacher feedback with metalanguage. Teachers therefore need to take caution about what kind of metalanguage to use and when to use it.

Implications for Research

Concerning the implications for research, I outline two issues below: uptake and teacher-feedback-related comments. Both are discussed in terms of their validity and/or reliability as a measure of noticing and learning.

Uptake needs to be supported by other measures

One thing that stood out throughout the previous three chapters is that the same category of a variable (the characteristic of feedback episodes, nonlinguistic cues, or teachers' metalanguage) sometimes had totally different effects on the different noticing and learning measures. For example, while indirect episodes resulted in a significantly higher level of uptake and successful uptake, they led to a significantly lower level of teacher-feedback-related comments and correct answers in the immediate test. The rates of uptake and successful uptake have been found to run counter to teacher-feedback-related comments and test results. This raises a serious issue. Although uptake has been reiterated by many researchers (Ellis et al., 2001a; Loewen, 2004; Sheen, 2006) to be an optional move, it has been widely adopted in different feedback studies to examine the effectiveness of feedback. The tables below show the results of chi-square analysis on the rate of teacher-feedback-related comments and test results in relation to the occurrence and successfulness of uptake. The p value in all cases are >.05, indicating that the occurrence and successfulness of uptake did not predict the rate of teacher-feedback-related comments or test results. This finding suggests that uptake and successful uptake cannot always serve as valid measures for noticing and learning, they therefore should be supported and supplemented by other measures.

Table 88 Uptake and teacher-feedback-related comments

	Total number of comments	Teacher-feedback-related	Non-teacher-feedback-related	Other
Uptake	182	130 (71.4%)	20 (11.0%)	32 (17.6%)
No uptake	60	43 (71.7%)	7 (11.7%)	10 (16.7%)
No opportunity	37	28 (75.7%)	3 (8.1%)	6 (16.2%)

$X^2(4, n=279)=.421, p=.981$

Table 89 Successfulness of uptake and teacher-feedback-related comments

	Total number of comments	Teacher-feedback-related	Non-teacher-feedback-related	Other
Successful	79	59 (74.7%)	7 (8.9%)	13 (16.5%)
Acknowledge	67	48 (71.6%)	8 (11.9%)	11 (16.4%)
Inconclusive	34	21 (61.8%)	5 (14.7%)	8 (23.5%)

$X^2(4, n=180)=2.139, p=.710$ (Unsuccessful uptake and partially successful uptake were excluded due to low expected cell counts)

Table 90 Uptake and test results

	Immediate test			Delayed test		
	Number of test items	Correct	Incorrect	Number of test items	Correct	Incorrect
Uptake	162	108 (66.7%)	54 (33.3%)	172	90 (52.3%)	82 (47.7%)
No uptake	82	54 (65.9%)	28 (34.1%)	62	31 (50%)	31 (50%)
No opportunity	28	14 (50%)	14 (50%)	48	29 (60.4%)	19 (39.6%)
	$X^2(2, n=272)=2.972, p=.226$			$X^2(2, n=282)=1.312, p=.519$		

Table 91 Successfulness of uptake and test results

	Immediate test			Delayed test		
	Number of test items	Correct	Incorrect	Number of test items	Correct	Incorrect
Successful	75	57 (76.0%)	18 (24.0%)	92	52 (56.5%)	40 (43.5%)
Acknowledge	60	36 (60.0%)	24 (40.0%)	58	31 (53.4%)	27 (46.6%)
Inconclusive	25	15 (60.0%)	10 (40.0%)	21	7 (33.3%)	14 (66.7%)
	$X^2(2, n=160)=4.650, p=.098$			$X^2(2, n=171)=3.711, p=.156$		

(Unsuccessful uptake and partially successful uptake were excluded due to low expected cell counts)

Teacher-feedback-related comments can be a reliable data source

Another thing that stood out in the findings of this study is the relationship between teacher-feedback-related comments and test results. Since the stimulated recall stimuli

and test items were developed from different observations, it is impossible to examine the relationship between the two with chi-square analysis. The available information, however, shows that the rate of teacher-feedback-related comments in general corresponds well with test results. Although there are some exceptions, they are not frequent. Given the relationship between noticing and learning, this suggests that teacher-feedback-related comments could be a reliable predictor of noticing and learning. In general, when there are more teacher-feedback-related comments from a particular type of feedback, there are also more correct answers for items developed from such episodes in the immediate test and delayed test. One typical example is the presence of general extralinguistic cues. That is, episodes with general extralinguistic cues led to significantly more teacher-feedback-related comments and learners provided significantly more correct answers for items developed from such episodes in the two tests. This result means that extralinguistic cues predicted both more noticing and more learning.

7.3 Reflection on the present study

This section is a reflection on the design of the present study. The study is flawed in several ways. First of all, many during-observation factors that could have influenced the results were not operationalized. For example, certain types of feedback episodes occurred frequently in some classes but rarely in other classes. Some teachers organized more communicative activities in their lessons than other teachers. Learners' proficiency levels, which have been found to affect noticing and learning (Mackey & Philp, 1998; Philp, 2003), were different from class to class. There were also big differences among learners in the same class. For example, in some classes a few learners were always

interacting with teachers while others were always quiet. All these inter- and intra-class variations could have played a role in the findings but not considered. Another thing the study should have taken into account is post-observation factors. Data collection took place in part of an intensive English program. Learners attended various classes between the observations and the tests or stimulated recall interviews. Most learners had to finish certain amount of homework. In a second language environment, they might also improve their English when socializing with native speakers. All these activities might have contributed to the observed effect of teacher feedback and yet not controlled or documented. A third drawback of the study is that the analysis is not deep enough. Inferential information was mostly drawn from chi-square tests and adjusted standardized residuals. Due to the complexity of categories involved as well as limited resources, no further analysis (e.g., regression) was made to better investigate how well the independent variables predicted the dependent variables. Moreover, chi-square analysis was not conducted with all variables due to its expected cell count requirement. A detailed analysis of the stimulated recall comments was made but left out because it did not directly answer the research questions. A fourth limitation of the study is the timing of the stimulated recall interviews. As a result of the research setting and again lack of resources, the stimulated recall interviews on average were conducted later than desired. This could have caused more loss of memory, reducing the veridicality of the noticing data. A fifth thing that could have strengthened the study would have been to have teacher interviews. Some studies (e.g., Mackey et al., 2007) have found that teachers' views on giving feedback can help reveal the relationship between teachers' intentions and learners' interpretations of the feedback. Due to the lack of data on teacher views, the

interpretation of the results in many cases had to draw on assumptions without enough evidence. The addition of teacher interviews to the research protocol would have helped to solve this problem.

Despite the various drawbacks, the study is strong and unique in many ways too. First of all, all the data were based on natural classroom interactions. Because many factors (e.g., the type of lessons and activities) were beyond my control, the coding and data analysis might not be as clean as that of laboratory studies. However, the data do show the interactions in real life ESL classrooms. The findings therefore allow for a different perspective on interaction, and they add some nuance to research using laboratory settings.

Second, unlike other studies that only operationalized noticing with a single measure, for example, uptake charts or think-aloud protocols, this study triangulated the measurement of noticing with both stimulated recall comments and learner uptake. The triangulation of two measures, one offline and one online, might have helped to elicit more thorough noticing data.

Third, studies which used stimulated recall interviews to elicit noticing data were often conducted in an experimental setting. In some studies, the stimuli contained specified language structures; in some, teachers were told what kinds of feedback types were desired. In the current study, the stimuli for the recall interviews were totally based on natural classroom interactions. The teachers were not asked to teach any particular

language structure or use any particular feedback type just for the purpose of the study. The stimuli used for the recall interviews were therefore more classroom-based and the noticing data elicited from these stimuli might better reflect learners' cognitive processes in real time classroom interactions.

Fourth, apart from factors that have received heavy attention but still call for further investigation – the characteristics of feedback, this study also investigated factors that have received little attention – nonlinguistic cues and teachers' metalanguage use. The inclusion of these factors aligned with and added to the findings of the few researchers who have worked on them.

Finally, the use of video-recordings made a difference. Most, if not all, studies which investigated teacher feedback and learner uptake used only audio-recordings of classroom interactions. In this study, I used a video camera to record classroom interactions. The videotapes helped me to capture the minute detail that might otherwise have slipped away. Consequently, I was able to examine learners' responses to teacher feedback with gestures and other body language, which few researchers have looked at.

7.4 Suggestions for future research

Based on what has been discussed in the previous sections, some tentative suggestions can be made about future research on the noticing and effectiveness of teacher feedback. These are listed below, without any more redundant discussion:

1. Nonlinguistic cues play an important role in the noticing and effect of teacher feedback and therefore deserve more attention. It would help to have a study in which nonlinguistic cues are the primary focus of data collection and in which a well established taxonomy is used to analyze the multiple effects of nonlinguistic cues.
2. Learners' nonverbal responses to teacher feedback can make a difference to research findings and therefore should be considered in feedback research. We need a study that examines how learners' nonverbal responses contribute to the dynamics of teacher-student interactions, and how such responses help with the negotiation of meaning and subsequent learning.
3. When a study involves classroom observations, post-observation learning opportunities could affect learners' language development. Therefore, we need a study that measures post-observation learning opportunities at various intervals and in various language use situations.
4. As I mentioned, the inclusion of both teacher views and learner views can help researchers better interpret learners' perceptions of teacher feedback. It would be a good idea to conduct a study similar to mine (but not necessarily identical) and add the component of teacher interviews.

7.5 Concluding remarks

In conclusion, the findings of the present study indicate that all the three factors examined -- the characteristics of feedback episodes, nonlinguistic cues in teacher feedback, and teachers' use of metalanguage -- can affect learners' noticing of teacher feedback and their language improvement. Although the findings may not apply to every

teaching context, they carry practical implications for classroom teaching to a certain degree. Moreover, while lending support to the findings of existing research, the study also raised important methodological issues. One finding is that learner uptake alone is not a sufficient measure for the noticing and effect of teacher feedback. Therefore, it is necessary to triangulate multiple measures in order to increase the validity of research data. On the other hand, teacher-feedback-related comments from stimulated recall interviews that were based on natural classroom interactions could be a reliable predictor of noticing and learning. Such interviews therefore can be a valuable tool to collect noticing and learning data.

REFERENCES

Adams, R. (2003). L2 output, reformulation, and noticing: Implications for IL development. *Language Teaching Research*, 7, 347-376.

Alanen, R. (1995). Input enhancement and rule presentation in second language acquisition. In R. Schmidt (Ed.), *Attention and Awareness in Foreign Language Learning and Teaching (Technical Report #9)* (pp. 259-302). Honolulu, Hawai'i: University of Hawai'i, Second Language Teaching and Curriculum Center.

Allwright, B. (1984). Why don't learners learn what teachers teach? The interaction hypothesis. In D. Singleton & D. Little (Eds.), *Language Learning in Formal and Informal Contexts* (pp. 3-18). Dublin: Irish Association for Applied Linguistics.

Ano, K. (1998). A study of the output hypothesis: Cognitive processes of speaking a foreign language. *Journal of Japan-Korea Association of Applied Linguistics, 2*, 175-204.

Ayoun, D. (2001). The role of negative and positive feedback in the second language acquisition of the passé composé and imparfait. *The Modern Language Journal, 85*(2), 226-243.

Baker, C. L. (1979). Syntactic theory and the projection problem. *Linguistic Inquiry, 10*, 533-581.

Basturkmen, H., Loewen, S., & Ellis, R. (2002). Metalanguage in focus on form in the communicative classroom. *Language Awareness, 11*(1), 1-13.

Berman, R. A. (1979). Rule of grammar or rule of thumb? *International Review of Applied Linguistics, 17*(4), 279-302.

Birdsong, D. (1989). *Metalinguistic Performance and Interlinguistic Competence*. New York: Springer.

Bley-Vroman, R. (1989). What is the logical problem of foreign language learning? In S. Gass & J. Schachter (Eds.), *Linguistic Perspectives on Second Language Acquisition* (pp. 41-68). Cambridge, England: Cambridge University Press.

Block, D. (2003). *The Social Turn in Second Language Acquisition*. Edinburgh: EUP.

Borg, S. (1998). Talking about grammar in the foreign language classroom. *Language Awareness, 7*(4), 159-175.

Borg, S. (1999). Teacher's use of grammatical terminology in the second language classrooms: A qualitative study. *Applied Linguistics, 20*, 95-126.

Brown, R. & Hanlon, C. (1970). Derivational complexity and the order of acquisition in child speech. In J. Heyes (Ed.), *Cognition and the Development of Language* (pp.155-207). New York: Wiley.

Carpenter, H., Jeon, K. S., McGregor, D., & Mackey, A. (2006). Learners' interpretations of recasts. *Studies in Second Language Acquisition, 28*, 209-236.

Chomsky, N. (1975). *Reflections on Language*. New York: Pantheon.

Corder, S. (1967). The significance of learners' errors. *International Review of Applied Linguistics, 5*, 161-170.

Corder, S. (1973). *Introducing Applied Linguistics*. Harmondsworth: Penguin.

Davies, M. (2006). Paralinguistic focus on form. *TESOL Quarterly, 40*(4), 841-855.

Egi, T. (2004). Verbal reports, noticing, and SLA research. *Language Awareness, 13*, 243-264.

Egi, T. (2007a). Recasts, learners' interpretations, and L2 development. In A. Mackey (Ed.), *Conversational Interaction in Second Language Acquisition: A Series of Empirical Studies* (pp. 249-267). Oxford: Oxford University Press.

Egi, T. (2007b). Interpreting recasts as linguistic evidence: The roles of linguistic target, length, and degree of change. *Studies in Second Language Acquisition, 29*(4), 511-537.

Egi, T. (2008). Investigating stimulated recall as a cognitive measure: Reactivity and verbal reports in SLA research methodology. *Language Awareness, 00*(0), 1-17.

Ellis, R. (2005). Measuring implicit and explicit knowledge of a second language: A psychometric study. *Studies in Second Language Acquisition, 27*, 141-172.

Ellis, R., Basturkmen, H., & Loewen, S. (2001a). Learner uptake in communicative ESL lessons. *Language Learning, 51*(2), 281-318.

Ellis, R., Basturkmen, H., & Loewen, S. (2001b). Preemptive focus on form in the ESL classroom. *TESOL Quarterly, 35*(3), 407-432.

Ellis, R. & Sheen, Y. (2006). Reexamining the role of recasts in second language acquisition. *Studies in Second Language Acquisition, 28*, 575-600.

Ellis, R., Tanaka, Y., & Yamazaki, A. (1994). Classroom interaction, comprehension and

the acquisition of L2 word meanings. *Language Learning, 44*, 449-491.

Ericsson, K. & Simon, H. (1993). *Protocol Analysis: Verbal Reports as Data* (2nd ed.). Boston: MIT Press.

Faech, C. (1985). Meta talk in FL classroom discourse. *Studies in Second Language Acquisition, 7*(2), 184-199.

Faraco, M., & Kida, T. (2008). Gesture and the negotiation of meaning in a second language classroom. In S. G. McCafferty & G. Stam (Eds.), *Gesture: Second Language Acquisition and Classroom Research* (pp. 280–297). London: Routledge.

Farrar, M. J. (1992). Negative evidence and grammatical morpheme acquisition. *Developmental Psychology, 28*, 90-98.

Gass, S. (1988). Integrating research areas: A framework for second language studies. *Applied Linguistic, 9*, 198-217.

Gass, S. (1997). *Input, Interaction, and the Second Language Learner*. Mahwah, NJ: Lawrence Erlbaum Associates.

Gass, S. (2003). Input and interaction. In C. Doughty & M. Long (Eds.), *The Handbook of Second Language Acquisition* (pp. 224-255). Malden, MA: Blackwell.

Gass, S. & Mackey, A. (2000). *Stimulated Recall Methodology in Second Language Research*. Mahwah, NJ: Lawrence Erlbaum Associates.

Gass, S. & Mackey, A. (2006). Input, interaction and output. *AILA Review, 19*, 3-17.

Gass, S. & Mackey, A. (2007). *Data Elicitation for Second and Foreign Language Research*. Mahwah, NJ: Lawrence Erlbaum Associates.

Gass, S. & Selinker, L. (2001). *Second Language Acquisition: An Introductory Course* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Gass, S. & Veronis, E. (1989). Incorporated repairs in NNS discourse. In M. Einstein (Ed.), *Variation and Second Language Acquisition* (pp. 71-86). New York: Plenum.

Gullberg, M. & McCafferty, S. G. (2008). Introduction to gesture and SLA: Toward an integrated approach. *Studies in Second Language Acquisition, 30*, 133-146.

Halliwell, S. (1993). *Grammar Matters*. London: CILT.

Han, Z. (2002). A study of the impact of recasts on tense consistency in L2 output.

TESOL Quarterly, 36(4), 543-572.

Hawkins, B. (1985). Is an "appropriate response" always so appropriate? In S. Gass & C. Madden (Eds.), *Input in Second Language Acquisition* (pp. 162-178). Rowley, MA: Newbury House.

Hinkle, D. E., Wiersma, W., & Jurs, S. G. (2003). *Applied Statistics for the Behavioral Sciences* (5th ed.). Boston, MA: Houghton Mifflin Company.

Iwashita, N. (2003). Negative feedback and positive evidence in task-based interaction. *Studies in Second Language Acquisition, 25*, 1-36.

James, C. (1999). Language awareness: Implications for the language curriculum. *Language, Culture and Curriculum, 12*(1), 94-115.

Jenkins, S. & Parra, I. (2003). Multiple layers of meaning in an oral proficiency test: The complementary roles of nonverbal, paralinguistic, and verbal behaviors in assessment decisions. *The Modern Language Journal, 87*(i), 90-107.

Jourdenais, R. (2001). Protocol analysis and SLA. In P. Robinson (Ed.), *Cognition and Second Language Instruction* (pp. 354-375). Cambridge: Cambridge University Press.

Kida, T. (2008). Does gesture aid discourse comprehension in the L2? In S. G. McCafferty & G. Stam (Eds.), *Gesture: Second Language Acquisition and Classroom Research* (pp. 131-156). London: Routledge.

Kim, H. (1995). Intake from the speech stream: Speech elements that learners attend to. In R. Schmidt (Ed.), *Attention and Awareness in Foreign Language Learning and Teaching (Technical Report #9)* (pp. 65-84). Honolulu, Hawai'i: University of Hawai'i, Second Language Teaching and Curriculum Center.

Krashen, S. (1982). *Principles and Practice in Second Language Acquisition*. London: Pergamon.

Krashen, S. (1985). *The Input Hypothesis: Issues and Implications*. Oxford: Pergamon Press.

Kucer, S. B. (2005). *Dimensions of Literacy: A Conceptual Base for Teaching Reading and Writing in School Settings* (2nd ed.). Mahwah, N.J.: Lawrence Erlbaum Associates.

Leow, R. (1997). Attention, awareness and foreign language behavior. *Language Learning, 47*, 467-505.

Leow, R. (2000). A study of the role of awareness in foreign language behavior: Aware

vs. unaware learners. *Studies in Second Language Acquisition, 22*, 557-584.

Leow, R. (2002). Models, attention and awareness in SLA: A response to Simard and Wong's 'Alertness, orientation, and detection: The conceptualization of attentional functions in SLA.' *Studies in Second Language Acquisition, 24*, 113-119.

Lightbown, P. (1998). The importance of timing in focus on form. In C. Doughty & J. Williams (Eds.), *Focus on Form in Classroom Second Language Acquisition* (pp. 177-196). Rowley, MA: Newbury House.

Loewen, S. (2002). The occurrence and effectiveness of incidental focus on form in meaning-focused ESL lessons. Unpublished Doctoral Thesis, The University of Auckland, Auckland, New Zealand.

Loewen, S. (2003). Variation in the frequency and characteristics of incidental focus on form. *Language Teaching Research, 7*(3), 315-345.

Loewen, S. (2004). Uptake in incidental focus on form in meaning-based ESL lessons. *Language Learning, 54*(1), 153-188.

Loewen, S. (2005). Incidental focus on form and second language learning. *Studies in Second Language Acquisition, 27*, 361-386.

Loewen, S. & Philp, J. (2006). Recasts in the adult English L2 classroom: Characteristics, explicitness, and effectiveness. *The Modern Language Journal, 90*(iv), 536-556.

Long, M. (1983). Linguistic and conversational adjustments to non-native speakers. *Studies in Second Language Acquisition, 5*, 177-194.

Long, M. (1985). Input and second language acquisition theory. In S. Gass & C. G. Madden (Eds.), *Input in Second Language Acquisition* (pp. 377-393). Rowley, MA: Newbury House.

Long, M. (1991). Focus on form: A design feature in language teaching methodology. In K. De Bot, R. Ginsberg, & C. Kramsch (Eds.), *Foreign Language Research in Crosscultural Perspective* (pp. 39-52). Amsterdam: Johns Benjamins.

Long, M. (1996). The role of the linguistic environment in second language acquisition. In W. C. Ritchie & T. K. Bhatia (Eds.), *Handbook of Second Language Acquisition* (pp. 413-468). San Diego, CA: Academic Press.

Long, M. (2007). *Problems in SLA*. Mahwah, NJ: Lawrence Erlbaum Associates.

Long, M. & Robinson P. (1998). Focus on form: Theory, research and practice. In C.

Doughty & J. Williams (Eds.), *Focus on Form in Classroom Second Language Acquisition* (pp. 15-41). Cambridge: Cambridge University Press.

Lyster, R. (1998a). Negotiation of form, recasts, and explicit correction in relation to error types and learner repair in immersion classrooms. *Language Learning, 48*(2), 183-218.

Lyster, R. (1998b). Recasts, repetition, and ambiguity in L2 classroom discourse. *Studies in Second Language Acquisition, 20*, 51-81

Lyster, R. (2004). Differential effects of prompts and recasts in form-focused instruction. *Studies in Second Language Acquisition, 26*, 399-432.

Lyster, R., & Mori, H. (2006). Interactional feedback and instructional counterbalance. *Studies in Second Language Acquisition, 28*, 269-300.

Lyster, R. & Ranta, L. (1997). Corrective feedback and learner uptake: Negotiation of form in communicative classrooms. *Studies in Second Language Acquisition, 20*, 37-66.

Mackey, A. (1999). Input, interaction, and second language development: An empirical study of question formation in ESL. *Studies in Second Language Acquisition, 21*, 557-587.

Mackey, A. (2006). Feedback, noticing and instructed second language learning. *Applied Linguistics, 27*(3), 405-430.

Mackey, A., Al-Khalil, M., Atanassova, G., Hama, M., Logan-Terry, A., & Nakatsukasa, K. (2007). Teachers' intentions and learners' perceptions about corrective feedback in the L2 classroom. *Innovation in Language Learning and Teaching, 1*(1), 129-152.

Mackey, A., Gass, S., & McDonough, K. (2000). How do learners perceive interactional feedback? *Studies in Second Language Acquisition, 22*, 471-497.

Mackey, A., McDonough, K., Fujii, A., & Tatsumi, T. (2001). Investigating learners' reports about the L2 classroom. *International Review of Applied Linguistics, 39*, 285-308.

Mackey, A. & Philp, J. (1998). Conversational interaction and second language development: Recasts, responses, and red herrings? *The Modern Language Journal, 82*(3), 338-356.

Mackey, A., Philp, J., Egi, T., Fujii, A., & Tatsumi, T. (2002). Individual differences in

working memory, noticing of interactional feedback and L2 development. In P. Robinson (Ed.), *Individual Differences and Instructed Language Learning* (pp. 181-209). Philadelphia: Benjamins.

Mackey, A. & Polio, C. (2009). Introduction. In A. Mackey & C. Polio (Eds.), *Multiple Perspectives on Interaction: Second Language Research in Honor of Susan M. Gass* (pp. 1-10). New York: Routledge.

Martin, H. R. (1981). The prosodic components of speech melody. *The Quarterly Journal of Speech, 67*, 81-92.

McDonough, K. (2005). Identifying the impact of negative feedback and learners' responses on ESL question development. *Studies in Second Language Acquisition, 27*(1), 79-103.

McDonough, K. & Mackey, A. (2006). Responses to recasts: Repetitions, primed production, and linguistic development. *Language Learning, 56*(4), 693-720.

McNeil, D. (1992). *Hand and Mind: What Gestures Reveal about Thought?* Chicago: The University of Chicago Press.

McNeil, D. (2005). *Gesture and Thought*. Chicago, IL: The University of Chicago Press.

Mehrebian, A. (1972). *Nonverbal Communication*. Chicago: Aldine-Atherton.

Nabei, T. & Swain, M. (2002). Learner awareness of recasts in classroom interaction: A case study of an adult EFL student's second language learning. *Language Awareness, 11*(1), 43-63.

Ohta, A. S. (2000). Re-thinking recasts: A learner-centered examination of corrective feedback in the Japanese language classroom. In J. K. Hall & L. Verplaeste (Eds.), *The Construction of Second and Foreign Language Learning through Classroom Interaction* (pp. 47-71). Mahwah, NJ: Erlbaum.

Oliver, R. (2000). Age differences in negotiation and feedback in classroom and pairwork. *Language Learning, 50*(1), 119-151.

Oliver, R. & Mackey, A. (2003). Interactional context and feedback in child ESL classrooms. *The Modern Language Journal, 87*(iv), 519-533.

Panova, I. & Lyster, R. (2002). Patterns of corrective feedback and uptake in an adult ESL classroom. *TESOL Quarterly, 36*(4), 573-595.

Philp, J. (2003). Constraints on "noticing the gap": Nonnative speakers' noticing of recasts in NS-NNS interaction. *Studies in Second Language Acquisition, 25*, 99-126.

Pica, T., Holliday, L., Lewis, N., & Morganthaler, L. (1989). Comprehensible output as an outcome of linguistic demands on the learner. *Studies in Second Language Acquisition, 11*, 63-90.

Polio, C. & Gass, S. (1998). The role of interaction in native speaker comprehension of nonnative speaker speech. *Modern Language Journal, 82*, 308-319.

Robinson, P. (1995). Attention, memory, and the "noticing" hypothesis. *Language Learning, 45*(2), 283-331.

Robinson, P. (1996). Learning simple and complex second language rules under implicit, incidental, rule search and instructed conditions. *Studies in Second Language Acquisition, 18*, 27-67.

Robinson, P. (1997). Individual differences and the fundamental similarity of implicit and explicit adult second language learning. *Language Learning, 47*, 45-99.

Robinson, P. (2003). Attention and memory during SLA. In C. Doughty & M. Long (Eds.), *The Handbook of Second Language Acquisition* (pp. 631-678). Malden, MA: Blackwell.

Rosa, E. & O'Neill, M. (1999). Explicitness, intake, and the issue of awareness: Another piece to the puzzle. *Studies in Second Language Acquisition, 21*, 511-556.

Schachter, J. (1988). Second language acquisition and its relationship to Universal Grammar. *Applied Linguistics, 9*, 219-235.

Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics, 11*(2), 129-158.

Schmidt, R. (1993). Awareness and second language acquisition. *Annual Review of Applied Linguistics, 13*, 206-226.

Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and Second Language Instruction* (pp. 3-32). Cambridge: Cambridge University Press.

Schmidt, R. & Frota, S. (1986). Developing basic conversational ability in a second language: A case study of an adult learner of Portuguese. In R. Day (Ed.), *Talking to Learn: Conversation in Second Language Acquisition* (pp. 237-322). Rowley, MA: Newbury House.

Sheen, Y. (2004). Corrective feedback and learner uptake in communicative classrooms across instruction settings. *Language Teaching Research, 8*(3), 263-300.

Sheen, Y. (2006). Exploring the relationship between characteristics of recasts and

learner uptake. *Language Teaching Research, 10*(4), 361-392.

Shriberg, E., Bates, R., Stolcke, A., Taylor, P., Jurafsky, D., Ries, K., Coccard, N., Martin, R., Meteer, M., & van Ess-Dykema, C. (1998). Can prosody aid the automatic classification of dialog acts in conversational speech? *Language and Speech, 41*(3-4), 443-492.

Sime D. (2006). What do learners make of teachers' gestures in the language classroom? *International Review of Applied Linguistics, 44*, 211-230.

Slimani, A. (1989). The role of topicalization in classroom language learning. *System, 17*, 223-234.

Song, S. (2007). Beginning ESL learners' noticing of morphological and syntactic changes in recasts. *Teachers College, Columbia University Working Papers in TESOL & Applied Linguistics, 7*(1), 1-25.

Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. Gass & C. Madden (Eds.), *Input in Second Language Acquisition* (pp. 235-253). Rowley, MA: Newbury House.

Swain, M. (1995). Three functions of output in second language learning. In G. Cook & B. Seidlhofer (Eds.), *Principle and Practice in Applied Linguistics: Studies in Honour of H. G. Widdowson* (pp. 125-144). Oxford: Oxford University Press.

Swain, M. (1997). The output hypothesis, focus on form and second language learning. In V. Berry, B. Adamson, & W. Littlewood (Eds.), *Applying Linguistics: Insights into Language in Education* (pp. 1-21). Hong Kong: English Centre, University of Hong Kong.

Swain, M. (2005). The output hypothesis: Theory and research. In E. Hinkel (Ed.), *Handbook of Research in Second Language Teaching and Learning* (pp. 471-484). Mahwah, NJ: Lawrence Erlbaum Associates.

Swain, M. & Lapkin, S. (1995). Problems in output and the cognitive processes they generate: A step toward second language learning. *Applied Linguistics, 16*, 371-391.

Tomlin, R. & Villa, V. (1994). Attention in cognitive science and SLA. *Studies in Second Language Acquisition, 16*, 185-204.

Trahey, M. (1996). Positive evidence in second language acquisition: Some long-term effects. *Second Language Research, 12*, 111-139.

Trahey, M. & White, L. (1993). Positive evidence and preemption in the second language

classroom. *Studies in Second Language Acquisition, 15*, 181-204.

VanPatten, B. (1994). Evaluating the role of consciousness in SLA: Terms, linguistic features, and research methodology. *AILA Review, 11*, 27-36.

Wagner-Gough, K. & Hatch, E. (1975). The importance of input in second language acquisition studies. *Language Learning, 25*, 297-308.

White, L. (1991). Adverb placement in second language acquisition: Some effects of positive and negative evidence in the classroom. *Second Language Research, 7*, 133-161.

Williams, J. (2001). The effectiveness of spontaneous attention to form. *System, 29*, 325-340.

Williams, J. (2005). Form-focused instruction. In E. Hinkel (Ed.), *Handbook of Research in Second Language Teaching and Learning* (pp. 671-691). Mahwah, NJ: Lawrence Erlbaum Associates.

MICHIGAN STATE UNIVERSITY LIBRARIES
3 1293 03063 1919