

**Recasts revisited:
The role of recasts in error detection and correction
by adult ESL students**

by

Laura Hawkes
B.A., Queen's University at Kingston, 2003

A Thesis Submitted in Partial Fulfillment of
the Requirements of the Degree of

MASTER OF ARTS

in the Department of Linguistics

© Laura Hawkes, 2007
University of Victoria

All rights reserved. This thesis may not be reproduced in whole or in part, by photocopy or other means, without the permission of the author.

SUPERVISORY COMMITTEE

**Recasts revisited:
The role of recasts in error detection and correction
by adult ESL students**

by

Laura Hawkes
B.A., Queen's University at Kingston, 2003

Supervisory committee

Dr. Hossein Nassaji (Department of Linguistics)

Supervisor

Dr. Hua Lin (Department of Linguistics)

Departmental Member

Dr. Catherine Caws (Department of French)

Additional Member

Dr. Emmanuel Hérique (Department of French)

External Member

ABSTRACT

Dr. Hossein Nassaji (Department of Linguistics)

Supervisor

Dr. Hua Lin (Department of Linguistics)

Departmental Member

Dr. Catherine Caws (Department of French)

Additional Member

Dr. Emmanuel Hérique (Department of French)

External Member

The research presented in this thesis examined whether incidental recasts provided by a teacher in a small-group outside a classroom were beneficial to adult ESL learners. The 26 ESL learners who participated in the study were divided into seven small groups (3-5 students per group) and each group participated in an oral activity with a teacher. During the activity, which was videotaped, the teacher provided incidental and extensive recasts to half of each student's errors; the other half of the errors received no feedback. One day after the small-group activity, students were tested on their ability to detect and correct errors in their own speech. From the videotape, three types of episodes were identified for testing: episodes that involved an error by the student and a recast (error+recast) (n = 111), episodes that involved an error but no recast (error-recast) (n = 111), and episodes that involved no error in the student's speech (correct) (n = 111). Students' ability to detect and correct their errors in the three types of episodes was assessed using two types of tests: a stimulated correction test (a video-based computer test) and a written test. Students' reaction time on the error detection portion of the stimulated correction task was also measured.

The results of the study showed that the students benefited from the recasts. It was found that students were able to detect more errors in error+recast episodes than in error-recast episodes (though this difference did not reach statistical significance). They were also able to successfully and partially successfully correct more errors from error+recast episodes than from error-recast episodes, and this difference was statistically significant on the written test. The reaction time results also point towards a benefit from recasts, as students were able to complete the task (slightly) more quickly for error+recast episodes than for error-recast episodes. It was concluded from the results that recasts, even when provided extensively and in a non-dyadic context, can be of benefit to ESL students.

TABLE OF CONTENTS

SUPERVISORY COMMITTEE.....	ii
ABSTRACT.....	iii
TABLE OF CONTENTS	v
LIST OF TABLES	vii
LIST OF FIGURES.....	viii
ACKNOWLEDGMENTS	ix
Chapter 1 – Introduction	1
1.1 BACKGROUND	1
1.2 PURPOSE OF THE STUDY	3
1.3 SIGNIFICANCE OF THE RESEARCH PROBLEM.....	3
1.4. ORGANIZATION OF THE THESIS	4
Chapter Two – Related Literature	6
2.1 INTRODUCTION.....	6
2.2 DEFINITIONS OF TERMINOLOGY	7
2.3 INTEREST IN RECASTS	10
2.3.1 <i>Pedagogical Perspective</i>	11
2.3.2 <i>Theoretical Benefits</i>	12
2.4 EARLY UPTAKE STUDIES	14
2.4.1 <i>Observational Classroom Studies</i>	14
2.4.2 <i>Dyadic Studies</i>	17
2.4.3 <i>Issues with Uptake</i>	19
2.5 RECALL TECHNIQUE STUDIES	21
2.6 PRE-/POST-TEST AND INDIVIDUALIZED TEST STUDIES	22
2.6.1 <i>Dyadic Studies</i>	23
2.6.2 <i>Small-Group and Classroom Studies</i>	25
2.7 ISSUES WITH PREVIOUS RECAST STUDIES	33
2.8 PURPOSE OF THE STUDY	38
2.9 RESEARCH QUESTIONS	39
Chapter 3 – Methodology.....	40
3.1 DESIGN	40
3.2 PARTICIPANTS	40
3.3 PROCEDURE.....	43
3.3.1 <i>Day One: Treatment</i>	43
3.3.2 <i>Creation of Testing Materials</i>	48
3.3.3 <i>The Testing Procedure</i>	54
Chapter 4 – Data Analysis and Results	57
4.1 DATA ANALYSIS	57
4.1.1 <i>Interaction Data</i>	57
4.1.2 <i>Scoring the Stimulated Correction Task</i>	57
4.1.3 <i>Scoring of Written Test</i>	63
4.1.4 <i>Reaction Times</i>	64

4.2 ERROR DETECTION RESULTS	65
4.2.1 <i>Stimulated Correction Task</i>	65
4.2.2 <i>Written Test</i>	67
4.2.3 <i>Comparison Between Stimulated Correction Task and Written Test</i>	69
4.2.4 <i>Summary</i>	72
4.3 ERROR CORRECTION RESULTS	73
4.3.1 <i>Stimulated Correction Task</i>	73
4.3.2 <i>Written Test</i>	80
4.3.3 <i>Comparison Between Stimulated Correction and Written Test</i>	84
4.3.4 <i>Summary</i>	88
4.4 REACTION TIME RESULTS.....	90
4.5 SUMMARY OF RESULTS.....	92
Chapter 5 – Discussion and Conclusions	95
5.1 DISCUSSION OF RESULTS	95
5.2 CONCLUSIONS	107
5.3 PEDAGOGICAL IMPLICATIONS.....	108
5.4 CONTRIBUTIONS TO RESEARCH DESIGN	109
5.5 LIMITATIONS AND IMPLICATIONS FOR FUTURE RESEARCH.....	110
REFERENCES.....	115
APPENDIX A.....	121
APPENDIX B	122
APPENDIX C	123
APPENDIX D.....	124
APPENDIX E	126

LIST OF TABLES

<i>Table 2.1.</i> Summary of recast studies.....	31
<i>Table 3.1.</i> Small-group information.....	41
<i>Table 4.1.</i> Coding of nature of modification according to episode type.....	63
<i>Table 4.2.</i> Students' responses on the error detection portion of the stimulated correction task.....	65
<i>Table 4.3.</i> Students' responses on the error detection portion of the written test.....	67
<i>Table 4.4.</i> Gains/losses in number of "error" responses from the stimulated correction task to the written test.	70
<i>Table 4.5.</i> Changes in error detection responses from the stimulated correction task to the written test.....	72
<i>Table 4.6.</i> Error modification patterns (stimulated correction task).....	74
<i>Table 4.7.</i> Combined error modification patterns, +recast and –recast episodes (stimulated correction task).....	76
<i>Table 4.8.</i> Same as recast vs. different than recast successful and partially successful modifications on +recast episodes (stimulated correction task).....	78
<i>Table 4.9.</i> Error modification patterns (written test).....	80
<i>Table 4.10.</i> Combined error modification patterns, +recast and –recast episodes (written test).	82
<i>Table 4.11.</i> Same as recast vs. different than recast successful and partially successful modifications on +recast episodes (written test).	83
<i>Table 4.12.</i> Gains/losses in modifications from stimulated correction task to written test.	84
<i>Table 4.13.</i> Gains in successful and partial successful modifications for +recast episodes from the stimulated correction task to the written test.	87
<i>Table 4.14.</i> Reaction times in error detection on the stimulated correction task.....	91
<i>Table 4.15.</i> Mean reaction times by nature of modification for +recast and –recast episodes.....	92

LIST OF FIGURES

<i>Figure 4.1.</i> Students' responses on the error detection portion of the stimulated correction task.....	66
<i>Figure 4.2.</i> Students' responses on the error detection portion of the written test.	68
<i>Figure 4.3.</i> Gains/losses in number of "error" responses from the stimulated correction task to the written test.	70
<i>Figure 4.4.</i> Error modification patterns on stimulated correction task.	74
<i>Figure 4.5.</i> Combined error modification patterns for +recast and –recast episodes on the stimulated correction task.....	77
<i>Figure 4.6.</i> Same as recast vs. different than recast successful and partially successful modifications on +recast episodes (stimulated correction task).....	78
<i>Figure 4.7</i> Error modification patterns on written test.....	81
<i>Figure 4.8.</i> Combined error modification patterns for +recast and –recast episodes on the written test.	82
<i>Figure 4.9</i> Same as recast vs. different than recast successful and partially successful modifications on +recast episodes (written test).	83
<i>Figure 4.10.</i> Gains/losses in modifications from stimulated correction task to written test.	85
<i>Figure 4.11.</i> Gains in successful and partial successful modifications for +recast episodes from the stimulated correction task to the written test.	88

ACKNOWLEDGMENTS

I would like to acknowledge the many people whose involvement has been integral to the successful completion of this thesis. First, I would like to thank my supervisor, Dr. Hossein Nassaji, for his endless hours of help. His expertise and knowledge were invaluable throughout the process, and this thesis would not have been possible were it not for his continued patience, encouragement, and belief in my ideas. I would also like to thank my committee members, Dr. Hua Lin and Dr. Catherine Caws, for their time, suggestions, and general support. Also, thank you to Dr. Emmanuel Hérique, who kindly agreed to be my external examiner.

Many additional members of the University of Victoria community helped to make this thesis possible. Thank you to Maureen Kirby for securing classrooms for my data collection and for assisting me with the administrative aspects of the thesis. Thank you to Chris Coey for his technological expertise. Thank you to Dr. Joseph Parsons, who showed me the value of setting daily and weekly goals. Thank you to my classmates from Ling. 575 for their comments and suggestions on the design of my study. Thank you also to Abbey Bell and Carolyn Pytlyk for their performance in the demonstration video. I would also like to extend an important thank you to the teachers at the English Language Centre who allowed me to come into their classes and the ESL students who generously volunteered their time to participate in my study.

I would also like to acknowledge the financial support I have received during my M.A. programme. Thank you to the Social Sciences and Humanities Research Council of Canada (SSHRC) for the Canada Graduate Scholarship, Master's Scholarship (2005-06). Thank you to the University of Victoria for an M.A. Fellowship (2006-07) and a

President's Research Scholarship (2005-06). I would also like to acknowledge the funding I received from my supervisor through his SSHRC research grant. Finally, thank you to my supervisor, who employed me as a research assistant (2005-07), and the University of Victoria, which employed me as a teaching assistant (2006-07).

Finally, thank you also to my friends and family, who have supported me in the completion of this thesis. My friends have supported me by acting as a sounding board for ideas (Jun Tian), being incredibly encouraging and enthusiastic (Sharon Caldwell), and going for coffee when I needed a break from writing (Qian Wang). Thank you also to my family. Thank you to my husband for patiently listening to my ideas (and frustrations) and for coming to walk me home on those nights that my data collection/test preparation kept me at school until midnight. Finally, a big thank you to my parents, who encouraged me from the very beginning to follow my interests and passions.

There are several reasons why recasts have received so much attention from researchers. First, they have been found to occur more often than any other type of corrective feedback in natural L2 classrooms (Lyster & Ranta, 1997; Panova & Lyster, 2002). Second, there are several theoretical reasons for assuming that recasts may be beneficial to students. Namely, it has been proposed (and debated) that recasts provide positive evidence (See Nicholas, Lightbown, & Spada, 2001) and negative evidence (See Gass & Varonis, 1994; Leeman, 2003), increase the saliency of target forms (See Leeman, 2003; Loewen & Philp, 2006; Nassaji, 2007), and promote interaction (See Ellis & Sheen, 2006).

There is a long list of researchers who have undertaken studies examining recasts (see Ammar & Spada, 2006; Ellis, Loewen, & Erlam, 2006; Iwashita, 2003; Loewen & Philp, 2006; Lyster, 2004; Lyster & Ranta, 1997; Mackey & Philp, 1998; Nabei & Swain, 2002; Nassaji, 2007, forthcoming; Panova & Lyster, 2002 among others). However, these studies have produced varying results. While some have shown significant benefits of recasts to learners, others have suggested that other types of corrective feedback might be more effective than recasts. Such varying results have meant that researchers and teachers are still unsure about the extent to which recasts are beneficial to L2 learners.

In addition to varying results, concerns have been raised regarding some of the methodological aspects of previous studies. Specifically, people have questioned whether uptake and other measures of noticing can be used as valid measures of the beneficial role of recasts (see Loewen, 2005; Loewen & Philp, 2006; Mackey & Philp, 1998; Nicholas et al., 2001). There is also concern that studies that use pre-selected target linguistic forms and/or intensive recasts may increase learners' attention to recasts and

the forms targeted by the recasts more than would naturally occur in L2 classrooms (see Ellis & Sheen, 2006; Nicholas et al., 2001). Finally, many researchers have questioned the applicability of results from dyadic interaction (interaction between two people) studies to real L2 classroom situations (see Ellis & Sheen, 2006; Loewen, 2005; Loewen & Philp, 2006; Lyster, 2004). Due to these real concerns with previous studies, there is a need for studies that examine recasts while addressing these issues.

1.2 Purpose of the Study

The present study was designed to examine the benefit of incidental, extensive recasts to adult ESL students participating in an oral small-group activity. Through making comparisons across learners' performance on errors that did and did not receive recasts, the study examines how recasts affect students' (a) ability to detect errors in their own speech, (b) ability to correct errors in their own speech, and (c) speed of error detection.

1.3 Significance of the Research Problem

The present study was motivated by the debate in the field of SLA as to how beneficial recasts are for L2 students. Previous studies of recasts have found varying results, which has been frustrating for both researchers and teachers. Currently, the debate rages on as to whether recasts are beneficial, and if they are, to what extent and under what situations. As such, more research is needed in this area, and the present study addresses this need. More specifically, the present study will address the need for more

studies to examine incidental, extensive recasts in non-dyadic interaction. While previous studies that have examined intensive recasts have added to our knowledge of recasts, they provide little information about the effectiveness of extensive recasts (recasts that are provided in response to a wide range of types of errors), which is the type of recast that often occurs in natural L2 classrooms. In the same way, dyadic studies do not show how students may benefit from recasts when they are provided in larger groups of students. Thus, the results in the area of extensive recasts in non-dyadic interaction that the present study will provide will be especially valuable.

In addition to the value of the results of the present study, the study will also contribute to the field of SLA in terms of its innovative research design. The new measure of language learning employed in the present study, *stimulated correction*, was designed to overcome some of the issues that have been raised regarding language learning measures used in previous recast studies. In addition, stimulated correction is a versatile measure that could be extended to the study of other type of corrective feedback in the future. As such, the methodological contributions of the present study are significant.

1.4. Organization of the Thesis

The thesis is organized as follows: Chapter two reviews the literature on recasts. It explains the theoretical and pedagogical underpinnings of recasts, discusses issues and concerns with current methodology, and presents the rationale for the present study. Chapter three describes the methodology, which includes the design, participants, and procedure. Chapter four presents the approach to data analysis and the research findings.

The results are organized into three main sections: Error detection, error correction, and reaction time. Finally, chapter five summarizes and discusses the findings of the study; it concludes with the limitations of the study and directions for future research.

Chapter Two – Related Literature

2.1 Introduction

This chapter reviews the relevant literature and research in the area. Though research on recasts does not have a long history within the field of second language acquisition (SLA), one thing is certain: Recasts are currently of great interest to researchers. The bulk of this chapter is devoted to examining this wealth of past research on recasts and discussing the issues that still remain. Before reviewing the studies, some terminology that is frequently used in this area of research is defined and examples are provided. Following these definitions, in section 2.3, the underlying motivations for the present study, both pedagogical and theoretical, are outlined.

Sections 2.4 to 2.6 present a review of research that has focused on recasts. This research has taken on many forms, varying in the method used to examine the effectiveness of recasts, the type of interaction in which the recasts were provided, and the type of recast provided. Section 2.4 looks at early observational studies of recasts that examined uptake. Section 2.5 outlines studies that have employed unique measures of noticing in their study of recasts. Section 2.6 focuses on the most recent research in the area of recasts: Research involving pre-/post-tests and individualized post-tests. In section 2.7, issues of concern in the study of recasts are presented and discussed. Section 2.8 explains the purpose of the present study and proposed improvements in testing procedures. Finally, section 2.9 introduces the research questions of the present study.

2.2 Definitions of Terminology

The literature in the area of interaction research has employed numerous terms that have sometimes been defined differently by different researchers. Definitions of terms employed in this study are reviewed below for clarification.

(i) corrective feedback:

An interlocutor's (e.g. a teacher's) response to a non-native speaker's error. The purpose of corrective feedback is to correct and/or signal the incorrectness/incomprehensibility of the erroneous form.

Examples (from Panova & Lyster, 2002):

- (2.a) S1: Yes, I have to...to find the answer on...on the book also?
 T: **In the book, yes. Both...in the book.** (corrective feedback)
 S1: In the book.
- (2.b) T: In a fast food restaurant, how much do you tip?
 S1: No money.
 T: **What's the word?** (corrective feedback)
 S1: Five...four...
 T: **What's the word...in a fast food restaurant?**
 (corrective feedback)
 S2: Nothing.
 T: Nothing, yeah.

Corrective feedback has been categorized into different types by different researchers, but most have used some combination of the six types proposed by Lyster

and Ranta (1997): explicit correction, recasts, clarification requests, metalinguistic feedback, elicitation, and repetition.

(ii) recast:

A type of corrective feedback that is generally defined as “involv[ing] the teacher’s reformulation of all or part of a student’s utterance, minus the error” (Lyster & Ranta, 1997, p. 46). Although this general definition of recasts exists, there is a considerable amount of variation in researchers’ more specific definitions of recasts (see Ellis & Sheen, 2006; Nicholas et al., 2001 for more discussion). For example, Long (2006) added that a recast is the reformulation of the *immediately preceding* utterance and that the focus of the interaction should be on *meaning*.

Examples (from Ellis & Sheen, 2006):

- (2.c) T: When you were in school?
 S: Yes. I stand in the first row.
 T: **You stood in the first row?** (recast)
 S: Yes, in the first row, and sit, ah, sat the first row.
- (2.d) S: Korean is more faster.
 T: **Is faster.** (recast)
 S: Is faster than English.

(iii) Pre-selected target (linguistic) form:

A grammatical rule/form (for example, the past-tense –ed, question formation, grammatical gender, etc.) that is singled out for study prior to any testing or treatment taking place during the research study.

(iv) incidental recasts:

Recasts that occur spontaneously during L2 communicative-based interaction; they are not pre-planned, and generally speaking they do not involve a pre-selected target form.

(v) intensive recasts:

Recasts that are directed towards only one or more pre-selected target form(s). In studies that have employed intensive recasts, errors other than those involving the pre-selected target forms would not receive recasts.

(vi) extensive recasts:

Recasts that are directed towards a wide range of linguistic errors during L2 interaction. They may include recasts directed towards pronunciation, morpho-syntactic, lexical errors and/or use of L1.

(vii) uptake:

Uptake refers to the learners' immediate verbal response to the corrective feedback and has been used in various studies as a measure of the noticing of, and therefore effectiveness of, corrective feedback (e.g., Ellis et al., 2001, Loewen, 2005; Lyster & Ranta, 1997; Panova & Lyster, 2002).

Examples of uptake (from small-group interaction in the present study):

- (2.e) T: Okay, and what about Janet's education?
 S1: Uh, Janet from Mount Allison University.
 T: **So Janet graduated from?** (Recast)
 S1: Uh, graduated from. (uptake, successful)
- (2.f) S3: ... Three years teaching English to children in Japan. And native speaker of English.
 T: **So she is a native speaker of English?** (Recast)
 S3: Yeah, yeah. (uptake, unsuccessful)

Example of no uptake (taken from small-group interaction in the present study):

- (2.g) S3: I can't uh oh sometimes I can't understand.
 T: Mmmm.
 S1: really?
 S3: Especially movie.
 T: **Movies?** (Recast)
 S3: If I don't- If I English- British movie. (no uptake)

2.3 Interest in Recasts

While various types of corrective feedback have received increasing attention from researchers in recent years, no other type has received as much attention as recasts. As the literature review that follows demonstrates, recasts have been singled out for examination in a large number of studies. Before examining these studies and their results, I would like to outline the various pedagogical and theoretical motivations for studying recasts.

2.3.1 Pedagogical Perspective

The first reason for wanting to study recasts is that they frequently occur naturally in L2 classrooms. Early observation classroom studies on corrective feedback were carried out with the purpose of describing what types of corrective feedback are being provided and how often they are being provided. It quickly became apparent that corrective feedback types were not all equal in terms of how often they were provided by teachers. Specifically, these studies found that recasts occurred much more often than any other type of corrective feedback. This high percentage of recasts was found in French immersion classrooms (Lyster, 1998; Lyster & Ranta, 1997), ESL classrooms (Havranek, 1999; Panova & Lyster, 2002), and a German L2 classroom (Lochtman, 2002). While all of these studies found that recasts occurred more often than any other type of corrective feedback, the exact percentage of recasts varied greatly, between 30.5% and 72%. Iwashita (2003) and Nassaji (2007) also found that during NS-NNS dyadic interaction, more recasts were provided than any other type of feedback. From these results, it is clear that recasts hold a special place in teachers' tool belt of corrective feedback techniques. This fact that recasts are 'special' merits attention from researchers.

While it is difficult to conclude from observational studies alone why recasts might be used more often than other types of corrective feedback, a possible reason is that recasts are generally considered to be a form of implicit correction in that they do not interrupt the flow of meaningful interaction (see Doughty & Varela, 1998; Nicholas et al., 2001). Teachers who want to keep the primary focus of the lesson on meaning may see this as an advantage. In addition, teachers may feel that some other types of corrective

feedback, such as elicitation, clarification requests, and repetition, may slow down a lesson, as the teacher must wait for the student to retrieve the correct form.

2.3.2 Theoretical Benefits

There are numerous theoretical reasons for assuming corrective feedback, and recasts in particular, may be beneficial for language learning. Specifically, it has been proposed (and debated) that recasts provide positive and negative evidence, increase the saliency of target forms, and promote interaction.

First of all, recasts provide positive evidence since they supply the learner with the correct form (Ellis & Sheen, 2006; Nicholas et al., 2001). This sets recasts apart from some other types of corrective feedback that encourage the learner to supply the correct form (i.e. elicitation, clarification requests). Krashen (1982) firmly believed that positive evidence is sufficient for language learning. While Schmidt (1990, 1995) and other advocates of focus-on-form believe that negative evidence is also needed for language learning, no one would deny that language learning would be extremely difficult, if not impossible, without positive evidence.

In addition to positive evidence, many claim that recasts may be beneficial because they provide negative evidence. Negative evidence is defined as “information regarding the impossibility of certain linguistic structures in the language being acquired” (Leeman, 2003, p. 38). Recasts, it is proposed, signal to the learner that there is something wrong with what they have said, thus providing negative evidence. Schmidt (1990, 1995) claimed that negative evidence can help learners notice the difference between their own inter-language (IL) forms and the target-language (TL) forms. He

proposes that it is this “noticing the gap” that will push learners to make changes to their IL forms, thus moving closer to native-like speech on the IL continuum. Gass and Varonis (1994) have also claimed that “the perception of a gap or mismatch may lead to grammar restructuring” (p. 299). However, Ellis and Sheen (2006) point out that recasts can only be considered to provide negative evidence if learners interpret them as corrective in nature. In addition, even if recasts can provide negative evidence, it is possible that this negative evidence is not beneficial (or not as beneficial as some other aspects of recasts) in promoting L2 learning (Leeman, 2003). At this point in time, the possibility that recasts provide negative evidence to learners and that this negative evidence could be beneficial for L2 learning continues to motivate researchers to examine recasts.

It has been argued that, in addition to providing positive and possibly negative evidence, recasts may increase the perceptual saliency of the target form as the juxtaposition of the learner’s utterance and the teacher’s recast will highlight the error (Farrar, 1990; Saxton, 1997). This increased saliency may help the learner to notice the negative evidence more than they would if provided with negative evidence at some later time or in some other form (Leeman, 2003). It has also been suggested that certain characteristics, such as emphatic stress and rising intonation, may add to the saliency of recasts, increasing their benefit (Loewen & Philp, 2006; Nassaji, 2006, 2007).

Finally, it has been suggested that recasts promote interaction. Long’s interaction hypothesis states that in addition to input, participation in interaction is needed for second language learning to occur (Long, 1981, 1983, 1996). When a learner makes an error and an interlocutor provides a recast (or other type of corrective feedback), this feedback

initiates interaction between the learner and the provider of the recast. Closely related to the idea of interaction is Swain's (1985) concept of pushed output, which proposes that learners must be pushed to produce modified output in order for L2 learning to occur. It is argued that without corrective feedback, learners will not be pushed to change their output. The findings from McDonough (2005) support this idea, showing that modified output following negative feedback was a significant predictor of question development.

In sum, recasts first became of interest to researchers because they are so prevalent in L2 classrooms. Since this discovery, researchers have proposed many theoretical explanations for why recasts would be beneficial to L2 learning. Based on these theories, numerous studies have been carried out on recasts, both in classrooms and in dyadic interactions. These studies and their results are discussed in the next section, but as we will see, the current picture of the effectiveness of recasts is anything but clear. Perhaps Ellis et al. (2006) put it best when they said, "there are some doubts as to how effective recasts are in promoting learning as well as to what kind of learning and knowledge they cater" (p. 342).

2.4 Early Uptake Studies

2.4.1 Observational Classroom Studies

The first studies to be published in the area of corrective feedback were mainly conducted in natural second language classrooms and were observational in nature. The aim of such studies was often to simply describe what was occurring naturally in these classrooms.

Lyster and Ranta (1997) conducted an observational classroom study early on that has been cited in nearly every study on corrective feedback published thereafter. Their study examined the extent of feedback and uptake in four grade 4 French immersion classrooms. They found that 62% of students' erroneous utterances received some type of corrective feedback and recasts were the most common type of feedback provided (55%). Overall, 45.2% of feedback resulted in uptake; however, it was clarification requests that lead to the most uptake, with recasts leading to uptake only 31% of the time (18% repair and 13% needs repair). In their discussion, Lyster & Ranta propose that uptake, as an indication of noticing, may be important for language learning, thus suggesting that their results show recasts may not be as effective as some other types of feedback (such as clarification requests).

This use of uptake as a measure of the (potential) effectiveness of corrective feedback was not isolated to Lyster and Ranta's (1997) study. In the years following their study, a number of other studies examined corrective feedback, and recasts in particular, using similar methodology but finding varied results. Ellis et al.'s (2001) observational study of two adult ESL classrooms found the relatively high overall uptake rate of 73.9% across both reactive and preemptive form-focused episodes (FFE). It is difficult to directly compare the results from this study with Lyster and Ranta's (1997) results as they did not divide types of corrective feedback in similar manners, and Lyster and Ranta (1997) examined only reactive FFEs. However, the difference in overall uptake rates (73.9% vs. 45.2%) suggest that context (grade 4 French immersion vs. adult ESL) may play a role in uptake levels.

Lyster (1998) conducted an observational study that investigated type of corrective feedback, type of errors, and levels of uptake in four grade 4 and grade 5 French immersion classrooms. He found that 61% of student errors received some type of corrective feedback from the teacher. The greatest percentage of feedback following phonological and grammatical errors was recasts (64% and 72% respectively). Following lexical errors, negotiation of form was the most frequent type of feedback (55%). He also found that, for phonological errors, uptake occurred most often following recasts. For lexical and grammatical errors, uptake was most frequent following negotiation of form. He concluded that while recasts may not be as effective for lexical and grammatical errors, they may be effective for phonological errors. With the results of Ellis et al. (2001) pointing to the possible role of context, and Lyster (1998) showing varied results across different types of errors, the debate about under what conditions and for what linguistic forms recasts may (or may not) be effective was well underway.

Wanting to more easily compare the results of Lyster and Ranta (1997) with uptake rates in a different classroom context, Panova and Lyster (2002) conducted an observational study in an early-intermediate adult ESL classroom in Quebec. They found, like Lyster and Ranta (1997), that recasts were the most common type of corrective feedback (55%). Also similar to Lyster and Ranta (1997), they found that the recasts led to lower uptake rates (40%) and repair rates (13%) than most other types of feedback (the exceptions were explicit correction and translation). Panova and Lyster concluded that recasts may not be viewed as corrective in nature by the (low-level) students and that while they may provide students with positive evidence, they “may not be the most effective way of promoting negative evidence” (p. 591).

Yet another classroom context was examined by Lochtman (2002), who carried out an observational study of high school L2 German classrooms in Belgium. The study found high levels of corrective feedback in response to student errors (90%), with recasts being the most common type of feedback (30.5%), although elicitations were a very close second (30.2%). Although their overall uptake rates for recasts were higher than Lyster and Ranta (1997) and Panova and Lyster (2002), like these studies, they found that recasts resulted in less uptake (47.5%) than other types of feedback (ranged from 48% to 100%).

These early observational studies all showed that recasts occurred more frequently than other type of corrective feedback in natural L2 classrooms; this was found in French immersion classrooms (Lyster, 1998; Lyster & Ranta, 1997), ESL classrooms (Ellis et al., 2001; Panova & Lyster, 2002), and a German classroom (Lochtman, 2002). However, the uptake results of these studies showed that recasts may not be as effective as other types of feedback. The exception to this was Lyster (1998), who found that recasts may be quite effective for phonological errors (but less effective for other types of errors).

2.4.2 Dyadic Studies

Several studies have examined recasts in the context of dyadic interaction, Braidı (2002) conducted an observational study on the use of recasts in 10 English NS-NNS dyads. Braidı found that recasts were employed in response to 15.45% of NNS's erroneous utterances, which was much lower than recast rates found in classroom studies. Next, breaking the term "uptake" down further than did Lyster and Ranta (1997), Braidı found that successful incorporation occurred 9.5% of the time following recasts, while

unsuccessful incorporation occurred 7.3% of the time and agreement occurred 28.67% of the time. Braidı concluded that the learners' short-term use of the recasts suggested that "recasts are of some utility to learners" (p. 32).

Also employing dyadic interaction, Nassaji (2007) conducted a study that examined recasts and elicitations in more detail than previous studies. Nassaji's study involving 42 adult intermediate-level ESL students and two teachers showed that reformulation occurred more often than any other type of feedback, and the most common subtype of reformulation was embedded recast + prompt (rising intonation and/or added stress and/or words highlighting the error). Results also showed that, for both elicitations and recasts, those produced with more explicit prompts (involving both rising intonation and additional verbal prompts) lead to more learner repair than did feedback with more implicit prompts or no prompts. Of all the subtypes of reformulations, single (isolated) recasts + prompts led to successful repair more often than any other subtype of recasts, suggesting that this type of reformulation, which both isolates the error and emphasizes it, may make the error particularly salient to learners.

Dyadic uptake studies have examined recasts in more detail than observational studies, and have shown that recasts are of some benefit to learners (Braidı, 2002), and that they are of particular benefit when they are isolated (Nassaji, 2007) and include prompts (Nassaji, 2007). Nassaji (2007) also found that recasts with an interrogative intonation were more likely to lead to uptake, but this conflicts with Sheen's (2006) classroom findings, which showed that recasts with a declarative intonation were more likely to lead to uptake.

2.4.3 Issues with Uptake

While several studies have been carried out using uptake rates as the sole measure of the effectiveness of recasts, not everyone is convinced uptake is either an appropriate or accurate measure. Before discussing the possible problems with uptake, each of the theoretical underpinnings for assuming uptake would be an appropriate measure of the effectiveness of recasts will be discussed. The use of uptake is perhaps most often linked with Schmidt's noticing hypothesis (Schmidt 1990, 1995), which proposed that attention is necessary for language learning to occur. It is argued that it is impossible for a student to uptake without noticing the corrective feedback, and so when a student uptakes, it must mean that they noticed the correction. However, as Mackey and Philp (1998) point out, while uptake can signal the presence of noticing, no uptake does not necessarily indicate an absence of noticing. As such, uptake cannot be considered to be an accurate measure of (all) noticing. In addition, as Schmidt himself mentions (Schmidt, 1995), noticing is only the first step in the acquisition process. As such, methods designed to measure only noticing cannot give us the full picture of L2 learning. Unfortunately, the measurement of uptake gives us neither a clear picture of noticing, nor indicates actual language learning.

In addition to the noticing hypothesis, the notion of pushed output is often brought into the rational for measuring uptake. Swain (1985) commented that if the learners are pushed to produce the correct form following corrective feedback (thus producing repair), it will help in the automatization of their L2 production. In other words, the more times they produce the form correctly, the more it will become automatic. However, we must not forget that uptake is usually defined as the acknowledgement produced by the student *immediately* following the corrective feedback (Loewen, 2005; Lyster & Ranta, 1997;

Panova & Lyster, 2002; among others). This definition of uptake would exclude subsequent correct uses of the form by the student. However, it is possible that a student could receive the recast, not uptake, but at some later point practice the new form on his or her own, and thus have the same benefit in terms of output practice as a student who uptook. As such, uptake may not be as closely tied to language learning as first proposed.

Additional evidence for the failure of uptake to measure language learning has come from studies that have combined uptake with other measures of language learning. Mackey and Philp's (1998) study on question development in dyadic interaction found that the presence or absence of uptake following recasts did not correlate with higher or lower scores on post-tests. Loewen (2005) demonstrated similar findings; through a binary logistic regression analysis, he found that no uptake (as opposed to uptake) was slightly more likely to result in correct answers on individualized test questions that dealt with grammatical and lexical errors. Finally, Loewen and Philp (2006) found that those characteristics of recasts that were predictive of uptake and repair were different than those that were predictive of accurate scores on individualized post-tests. Loewen and Philp (2006) also suggest that uptake may not be equally meaningful following all types of corrective feedback. Specifically, whereas uptake following feedback types such as elicitation and clarification requests "demonstrates whether a learner has grasped the problem or not, this result is not necessarily true of uptake following recasts" (Loewen & Philp, 2006, p. 539). The results of these three studies suggest that uptake may not be an accurate measure of the effectiveness of corrective feedback in promoting L2 learning and that we should question the conclusions that uptake studies make on the effectiveness of recasts. However, in questioning the usefulness of uptake we should not completely

dismiss these early studies on corrective feedback as there is no doubt that these early studies have contributed a great deal to the field of SLA, especially in terms of what corrective feedback types occur naturally in L2 classrooms and what learners' responses to feedback moves are. Nonetheless, to truly examine the effects of recasts on L2 learning, researchers need to look beyond uptake.

2.5 Recall Technique Studies

To address the problems of uptake, some researchers have searched for more reliable measures of noticing. This led to a few studies involving various recall techniques. Philp (2003) used cued immediate recall to test NNSs' memory of recasts in NS-NNS dyadic interaction. The NS provided recasts following errors in question formation, and whenever the NSS heard the cue (two knocks), he/she was required to repeat what the NS had just said. This cued immediate recall was hoped to serve as an indication of some level of noticing. Results showed that learners were able to repeat 60-70% of the recasts, but that certain factors (length of recast, number of changes in the recast, and learner level) affected their ability to recall the recasts. Philp took these results to suggest that, at least in dyadic interaction, learners do notice a considerable amount of recasts. Philp also concluded that those recast characteristics that led to higher recall rates (shorter recasts with fewer changes) "may be of more benefit to learners" (p. 117).

Captenter, Jeon, MacGregor and Mackey (2006) also conducted a study concerning learners' interpretations of corrective feedback. Their study involved showing video clips of teachers providing either recasts or repetitions to adult ESL learners. Half of the learners viewed full clips, including the student's utterance that prompted the

recast/repetition by the teacher. The other half viewed edited clips in which the student's erroneous utterance had been removed. Learners were then asked to judge whether they thought each of the teachers' utterances was a recast or a repetition. Results showed that learners were more likely to identify a recast as such when the video clip included the student's utterance. Carpenter et al. concluded that the key factor in interpreting recasts as corrective in nature may lie in the differences between the learner's utterance and the teacher's utterance. Although their study did not comment on the effectiveness of recasts, the results suggest that recasts may provide negative evidence (in addition to positive evidence), a matter that has been of some debate (see Ellis & Sheen, 2006).

Studies employing recall techniques are useful in helping us understand how learners perceive recasts, and they may be more accurate at assessing learners' noticing of recasts than uptake studies. However, we must remember that neither the noticing of recasts nor the interpretation of recasts as corrective in nature can necessarily be equated with L2 learning. If we want to have a clearer picture of the effect of recasts on L2 learning, we must look beyond the notion of noticing.

2.6 Pre-/Post-Test and Individualized Test Studies

In the movement away from noticing as the sole measure of the effectiveness of recasts, a number of researchers have used pre-/post-test designs. With the introduction of pre-/post-tests also came the increased use of pre-selected linguistic target forms. The need for such targets is clear: It is nearly impossible to pre-test linguistic forms when you do not know on which forms the learner will produce errors and receive spontaneous recasts. If a target form is pre-selected, learners can be given a pre-test on that form,

participate in tasks designed to elicit that form, receive recasts when they make errors on that form, and then be given post-test(s).

2.6.1 Dyadic Studies

Among the dyadic recast studies involving pre-selected target forms, there were a number of studies focused on English question formation. Mackey and Philp (1998) conducted a study on the effect of intensive recasts on question development in dyadic interaction between five NS and 35 NNS of English. The students were divided into three groups: negotiated interaction, recast, and control. The two treatment groups were then further divided into “readies” (higher level students) and “unreadies” (lower level students) based on their performance in the first session (pre-test). After eight sessions (five treatment and three test), results showed that the “readies” in the recast group outperformed the “readies” in the negotiated interaction group. The same was not true for the “unreadies”. Mackey and Philp concluded from this that recasts may be effective if and only if the students are developmentally ready to receive the information in the recasts. McDonough and Mackey (2006) also conducted a dyadic interaction study examining question formation, but their study examined both the effect of recasts on adult EFL learners’ question development and the relationship between learners’ responses to recasts and subsequent question development. The results showed that recasts were significantly correlated with question development, but no such relationship was found between the immediate repetition of recasts (repair) and question development. There was, however, a significant relationship between primed production

(when the learner used the form provided in the recast to ask a new question) and question development.

Further dyadic studies in recasts were carried out by Iwashita (2003) and Leeman (2003). In Iwashita's (2003) study, 55 beginner-level Japanese students were placed into either the control group (14 students), which participated in free talking, or the treatment group (41 students), which participated in three communicative tasks designed to elicit the Japanese locative-initial construction and the te-verb form. The treatment group received whatever feedback the interlocutor deemed natural to provide. All students were tested on the target items using a pre-test, immediate post-test, and delayed post-test (1 week later). It was found that the treatment group significantly outperformed the control group on both posttests. As for the relative contribution of the various interactional moves, results differed based on the target form and the learner's score on the pretest. For the locative construction, simple models were effective for above-average learners, but actually led to lower scores for below-average students. For the te-verb construction, recasts positively influenced posttest scores for both levels of students, which contrasts with Mackey and Philp's (1998) findings that recasts were more effective for students who were developmentally ready. Iwashita concluded that there was clear evidence that recasts can be effective in promoting short-term learning of certain form(s) (te-verb), but that recasts may not be effective for other forms (locative).

Leeman (2003) examined what component/aspect of recasts might be most useful to learners. Seventy-two adult L2 learners of Spanish were divided into four treatment groups to participate in dyadic interaction with the researcher. They were tested on noun-adjective agreement using a pre-test, immediate post-test, and delayed post-test, which

involved oral picture-difference tasks. Leeman found that learners who had received either recasts (negative evidence and enhanced-saliency) or enhanced-saliency treatments outperformed those who had received either negative evidence alone or no increased saliency (control group). Leeman concluded that recasts can indeed be beneficial to students, even on forms with low communicative value. She further concluded that recasts may mainly be beneficial due to their increasing the saliency of the target form.

The results of the dyadic studies summarized above have all shown some benefit from recasts. However, there are some limitations on just how beneficial the recasts were in each of the studies. While Mackey and Philp (1998) found that recasts were only beneficial to students who were “ready”, Iwashita found that recasts were effective for both lower and higher level students. However, this was true for one linguistic form but not for another. Finally, Leeman (2003) suggested that recasts may be more effective (only) when they increase the saliency of target form.

2.6.2 Small-Group and Classroom Studies

A number of researchers have chosen to undertake recast studies in small-group and classroom contexts. In an attempt to mimic the way in which group activities are conducted in L2 classrooms, Han (2002) conducted a small-scale study using small groups of students. Eight adult ESL students were divided into two groups: recasts and no feedback to participate in eight small-group treatment sessions with the researcher. The pre-selected linguistic target was tense consistency. The pre-/post-/delayed post-tests consisted of written and oral narration tasks. Recasts were provided to the recast group during oral interaction when they were not being tense-consistent (i.e. if they started in

the past tense and then used a present tense verb it would be recast as a past tense verb). The results of the oral tests showed that overall, both groups tended to use more past tense as the experiment progressed. On both written post-tests, the recast group noticeably increased their tense consistency, enlarging the gap between the two groups. Han concluded from this that the recasts impacted the students in a positive way and led to the students' better control over tense consistency. Han also proposed that the recasts may have heightened learners' awareness of the need for tense consistency, as evidenced in the recasts group's increased number of self-corrections on tense errors. However, Han also admitted that with only eight students in the study, we cannot draw any firm conclusions.

A couple of pre/post-test studies involving pre-selected linguistic targets have also compared the effectiveness of recasts vs. prompts. Lyster (2004), in a study involving 179 French immersion grade five students, found that prompts seemed better than recasts or no feedback, (but only on improving performance on the written tasks, not the oral tasks), "confirming that recasts, when compared to other feedback options, are not necessarily the most effective type of feedback in communicatively oriented classrooms" (p. 428). Ammar and Spada (2006) examined recasts vs. prompts in three grade 6 intensive ESL classrooms. After one instructional session and 11 practice sessions, they found that both the class receiving intensive prompts and the class receiving intensive recasts significantly outperformed the control group on the post-tests. Similar to Lyster (2004), they also found prompts to be more beneficial than recasts, but only on the written tests and not on the oral tests. They also found that those students who had preformed better initially seemed to benefit equally from recasts and prompts, while

those students who performed more poorly on the pre-test benefited more from prompts than from recasts, confirming Mackey and Philp's (1998) results that lower-level students may not be able to benefit from recasts as much as higher level students. Ammar and Spada concluded that while prompts may be more effective than recasts, especially for lower-level learners, both prompts and recasts may be of some benefit to learners. Nonetheless, in interpreting the results of Lyster (2004) and Ammar and Spada (2006), it should be kept in mind that both prompts and recasts were used in combination with form-focused instruction (FFI), and thus it is not possible to draw any conclusions from these study about the effectiveness of recasts in the absence of FFI.

Ellis, Loewen and Erlam (2006) used a pre-/post-test design to compare the effects of recasts vs. metalinguistic feedback on implicit and explicit knowledge of the past-tense -ed. Three intact ESL classes each received a different treatment: recasts (implicit feedback), metalinguistic feedback (explicit feedback), or no feedback (control group). In hopes of testing the students' implicit knowledge levels, an oral imitation pre-/post-test was employed; it involved listening to, stating degree of agreement with, and repeating using correct grammar 36 belief statements. A written grammaticality judgment test was used to examine explicit knowledge. The results varied depending on the time of test and the type of test, and Ellis et al. to conclude that metalinguistic feedback was more effective than recasts in promoting higher scores on both implicit and explicit knowledge tests (especially on the delayed post-tests), at least for past-tense -ed. Their results also showed some (weak) benefit of recasts on the implicit test, but not on the explicit test.

While Han (2002), Ammar and Spada (2006), and Ellis et al. (2006), and Lyster (2004) all found some benefits for recasts, just as the dyadic studies did, the findings are

not consistent across all factors. Specifically, results varied depending on the type of test employed and the time of the test (immediate vs. delayed post-tests). In general, more benefit from recasts was found on written tests than on oral tests (but the opposite was found in Ellis et al, 2006), and written tests also seemed to show more benefit for prompts over recasts than oral tests did. Delayed post-tests showed more benefit from any type of feedback (recast, prompt, and metalinguistic) than immediate post-tests. In addition, Ammar and Spada (2006) found an effect involving learner-level, with recasts being more beneficial than prompts for higher-level learners only.

In addition to the classroom and dyadic pre-/post-test studies that have employed intensive recasts directed at pre-selected target linguistic forms, some researchers have chosen to carry out studies that examined the effects of extensive recasts. Ishida (2004) attempted to lessen the unnatural situation of providing intensive recasts by providing recasts “whenever [the researcher] felt the need to confirm the meaning of the message” (p. 340). This included recasts following errors in the target form (the Japanese aspectual form *-te i-(ru)*), as well as in other forms. In Ishida’s small-scale study, four L2 learners of Japanese met individually with the researcher over a series of eight 30-minute sessions (two pre-tests, four treatments, and two post-tests). Results showed a positive correlation between the number of recasts a learner received and their accurate use of *-te i-(ru)*. While cautious about the results as only four learners were involved, Ishida concluded that recasts can result in significant increases in students’ accurate use of the target form, and that this is possible even when recasts are not focused solely on that target form.

Ishida (2004) widened the range of errors to receive recasts, but Nassaji (2006) went even further by examining truly extensive recasts using an innovative pre-/post-test

design with no pre-selected target form. Each of 42 learners was asked to write a description based on a series of pictures. A native speaker (NS) then collected the written description and asked the NNS to orally describe the story, keeping it as close to their written story as possible. During this oral interaction, corrective feedback (including recasts and elicitations) were provided whenever the NS felt it was appropriate. After the oral interaction, the written story was returned to the student and he/she was asked to make corrections. A similar delayed post-test was given two weeks later. The post-tests showed that learners successfully corrected more of the errors that had received recasts than the errors that had received elicitations. Another finding of the study was that explicit forms of feedback led to more successful corrections than implicit forms. This was true for both recasts and elicitations, but the advantage of explicitness over implicitness was greater in recasts. Nassaji concluded that recasts may be more beneficial when they include prompts.

Seeing that there was a lack of classroom-based studies on incidentally and extensively provided corrective feedback, Loewen (2005) conducted a classroom study on the effectiveness of spontaneous focus-on-form (both reactive and preemptive) on L2 learning. Although his study did not single out recasts for examination, the data was reexamined in Loewen and Philp (2006) to examine recasts in detail. Loewen (2005) observed 17 hours of natural classroom interaction, which he used to create individualized immediate and delayed post-tests based on the focus-on-form episodes (FFE) (the immediate and delayed post-tests were based on different FFEs). The post-tests consisted of oral suppliance, correction, and pronunciation tasks and each student was only tested on the questions pertaining to the FFEs in which they participated. As the

feedback was spontaneous and the researchers did not know ahead of time which forms would receive feedback, no pre-tests could be administered. Instead, the initial errors made by the students served as a type of pre-test (i.e. if the student made the error, it indicated that their knowledge of that form was to some degree incomplete). Uptake was also measured. Results presented in Loewen and Philp (2006) showed that recasts were the most common type of corrective feedback and that learners had an accuracy rate of 53% on the immediate posttest and 50% on the delayed post-test. Even though these scores were lower than those for inform and elicitation on the immediate post-test (64.9% and 75% respectively) and in between inform and elicitation scores on the delayed post-test (48.5% and 59.4% respectively), Loewen and Philp found these results encouraging, especially because the recasts were incidental and extensive. However, they also caution that correct answers do not necessarily demonstrate the acquisition of new linguistic knowledge (it could be consolidation of knowledge they already had). An additional finding of the study was that the characteristics of recasts predictive of uptake were different than those predictive of correct test scores, thus strengthening the claim that we should not look to uptake as a measure of the effectiveness of corrective feedback.

Ishida (2004), Nassaji (2006) and Loewen and Philp (2006) have all examined the effectiveness of extensive recasts in promoting L2 learning, and they have all found some benefit from recasts. This suggests that recasts, even when provided incidentally and extensively, may be beneficial to L2 students, both in dyadic interaction and in classroom interaction. However, more research is needed in this area, especially in the area of non-dyadic interaction, before firm conclusions can be drawn.

Table 2.1 provides a summary of the recast studies discussed in this chapter. As can be seen, a large number of studies over the last ten years have focused on recasts, and these studies have varied both in their methodology (i.e. measure of L2 learning, feedback types examined, context, use of pre-selected target forms, and use of intensive or extensive recasts) and in their results. The majority of these studies have found some benefit from recasts, but it is clear that some studies have found more benefit than others. In addition, different studies have compared recasts to different types of feedback (prompts, metalinguistic feedback, elicitations, etc.) and have often found that recasts may not be as effective as these other types of feedback (but again, the results are varying). Thus we are left with many results, but few firm conclusions about the role of recasts in L2 learning.

Table 2.1. Summary of recast studies.

	Measure of L2 learning	Feedback types examined	Context	Pre-selected target form	Intensive/ extensive
Lyster & Ranta (1997)	uptake	naturally arising*	classroom		extensive
Ellis, Basturkmen, & Loewen (2001)	uptake	naturally arising	classroom		extensive
Lyster (1998)	uptake	naturally arising	classroom		extensive
Panova & Lyster (2002)	uptake	naturally arising	classroom		extensive
Lochtman (2002)	uptake	naturally arising	classroom		extensive
Braidi (2002)	uptake	naturally arising	dyads		extensive
Nassaji (2007)	uptake	recasts and elicitations	dyads		extensive
Philp (2003)	cued immediate recall	recasts	dyads	English questions	intensive

Carpenter, Jeon, McGregor, & Mackey (2006)	student perception of video clips	recasts vs. repetitions	individual (no interaction)		extensive
Mackey & Philp (1998)	pre-/post-tests	recasts	dyads	English questions	intensive
McDonough & Mackey (2006)	pre-/post-tests and uptake	recasts	dyads	English questions	intensive
Iwashita (2003)	pre-/post-tests	naturally arising	dyads	Japanese locative-initial construction & te-verb	extensive
Leeman (2003)	pre-/post-tests	recasts vs. negative evidence vs. enhanced saliency	dyads	Spanish noun-adj. agreement	intensive
Han (2002)	pre-/post-tests	recasts	small-group	tense consistency	intensive
Lyster (2004)	pre-/post-tests	recasts vs. prompts	classroom	French noun gender	intensive
Ammar & Spada (2006)	pre-/post-tests	recasts vs. prompts	classroom	3 rd person singular determiners	intensive
Ellis, Loewen, & Erlam (2006)	pre-/post-tests	recasts vs. metalinguistic feedback	classroom	past tense -ed	intensive
Ishida (2004)	pre-/post-tests	recasts	dyads	Japanese –te I-(ru)	extensive
Nassaji (2006)	pre-/post-tests	recasts vs. elicitations	dyads		extensive
Loewen & Philp (2006)	post- tests and uptake	naturally arising, with a focus on reporting recasts	classroom		extensive

* *Note.* In studies labeled as examining *naturally arising* feedback, no types of feedback were preselected for study and the researchers examined all feedback types that arose naturally during the interaction.

2.7 Issues with Previous Recast Studies

In addition to the variations in results found in these recast studies, there is the concern that some studies may provide results that do not allow us to draw conclusions on the benefit of recasts for actual (real-life) L2 learning. As already discussed, studies that have employed uptake or other measures of noticing as the sole measure of L2 learning (9 of the 20) may not tell us very much about how recasts benefit L2 learning. Furthermore, the pre-/post-test studies have also been the subject of some debate. Specifically, there are concerns over the naturalness of dyadic studies, the use of pre-selected target linguistic forms, and the practice of intensive recasts.

In dyadic interactions, learners receive individual attention from a NS over an extended period of time and often over several sessions (for example, eight 30-minute sessions in Ishida, 2004; seven 15-25 minute sessions in Mackey & Philp, 1998). This is unusual in typical classroom situations where there are many students in one class (for example, 25 students/class in Panova & Lyster, 2002; 4-12 students/class in Sheen, 2006). Although students may receive some one-on-one time with the teacher in the classroom, it is more likely that most of their time interacting with the teacher is done in small-group and whole-class situations (Nassaji & Hawkes, 2006). It is possible that students pay more attention (and conversely pay more attention to recasts and/or target forms) in dyadic interaction than they do in small-group and whole-class interaction. This idea would be consistent with Nicholas, Lightbown and Spada's (2001) claim that "the dyadic nature of laboratory interactions...may help learners recognize the interlocutor's feedback as corrective" (p. 749). Lyster (2004) has also pointed out that individual

attention from a NS can alter results; he suggested that one reason his study found no differences between the effectiveness of recasts vs. prompts vs. no feedback when combined with form-focused-instruction may have been that the students benefited from the individualized attention they received during the three testing sessions and that this benefit neutralized any benefit from feedback type. Ellis and Sheen (2006) have also emphasized that the results of dyadic studies are not necessarily translatable/applicable to classroom situations. Thus, caution should be used when the results of dyadic studies are interpreted; it should not be assumed that similar results would occur in a natural classroom context.

A second concern with a number of the recast studies is the use of pre-selected linguistic target forms. As mentioned, target form(s) were chosen in order to be able to give pre-tests. Unfortunately, since the pre-test, tasks during treatment sessions, and post-tests were all designed to elicit the target form, the students may have picked up on this and directed their attention to that form. Thus, students may pay more attention to feedback provided on that form, altering the results of the study.

A third, closely related concern involves the provision of intensive recasts for the target form(s). The majority of the studies involving pre-selected target forms provided intensive recasts in response to errors in the target form while ignoring all other types of errors. This may have produced different results than it would have if recasts had been provided extensively, as is often found in natural classroom situations. As observational studies have shown, recasts are frequently used in L2 classrooms (Lochtman, 2002; Lyster, 1998; Lyster & Ranta, 1997; Panova & Lyster, 2002). Through examples of recasts provided in these studies and other observational-based studies (Loewen & Philp,

2006; Nassaji & Hawkes, 2006; Sabbagh, 1995; Sheen, 2006), it is clear that recasts were provided in response to a wide range of errors in these natural classroom situations. This paints a very different picture than that created in studies that provide intensive recasts directed towards one or more targeted forms. In fact, Ellis and Sheen (2006) went so far as to say, “the claim that recasts are most effective when they are focused and intensive (i.e. directed repeatedly at a single linguistic feature) is of little practical significance to teachers” (p. 597). Thus, while studies involving intensive recasts directed towards pre-selected linguistic targets can certainly add to our knowledge of recasts, they do not necessarily advance our knowledge of the role of extensive recasts in actual (i.e. natural) classroom situations.

After taking all of the concerns with recast studies into consideration, the long list of recast studies summarized in table 2.1 looks less conclusive. In fact, of the 11 studies that employed (pre-)/post-tests, less than half were non-dyadic (5 of the 11). Of these five studies, all but one employed a pre-selected target linguistic form and intensive recasts (the exception is Loewen & Philp, 2006).

The final concerns I would like to mention involve the testing methodology used in Loewen and Philp (2006), the one study that managed to avoid all of the previous concerns. First, the number of test items each student received in Loewen and Philp (2006) was quite low (average of 1.5 items per student per test). Without sufficient test items per student, it is difficult to draw firm conclusions from the findings. In addition, there are several concerns with the type of test that was employed. One issue with this type of test is the fact that the students were allowed to listen to the sentence more than once, and they were given as long as they needed to answer. Han and Ellis (1998) point

out that if the tests “allow subjects time to plan their sentences carefully they may tap primarily explicit knowledge” (p. 3). As the error was initially made in oral real-time interaction, it can be assumed that it was made under time constraints and possibly under a stressful situation for the student. Unfortunately, the test procedure did not reflect this type of situation as much as it could have. Secondly, the test questions were taken out of context. As each question consisted of only one sentence, it might be difficult for the student to remember that they in fact produced that sentence, let alone remember the context in which they spoke it or their intended meaning. Without contextual clues, it can be difficult even for a native speaker to make grammaticality judgments. This may be particularly true for errors such as verb tense, choice of articles, etc. Thirdly, it was the tester, a native-speaker of English, who read the sentences aloud. It is possible that the students would be less likely to find errors in sentences that are read by a NS than those read by a NNS because students are not usually asked to find errors in the speech of a NS. In addition to these concerns with the test, there is the problem of what to compare the results of the post-test to. As it is incidental recasts that are being examined, it is not possible to give a pre-test similar in style to the individualized post-test. Therefore, how can we know whether 50% accuracy on post-tests is good or bad? We need to find something meaningful with which to compare it before we can make any claims about the beneficial role of recasts.

Finally, although not mentioned by previous researchers as an issue, I feel that we need to fine-tune the notion of “the effectiveness” or the “beneficial role” of recasts. Tests used in previous studies have measured students’ ability to correct erroneous forms, provide correct forms (i.e. fill-in-the-blanks), and/or produce correct forms in their

speech. However, I feel that all of these tests measure “all or nothing” language learning: Either the student has learned the form or he/she has not. While tests lacking any fine-tuned analysis may be fine when examining intensive recasts, they may miss any benefit from extensive recasts. For example, it is conceivable that in order to attain “full learning” (i.e. able to produce the target form), a learner needs to receive several recasts on that form, but that does not necessarily mean that “no learning” occurred after a single recast. It is possible that a single recast could push the learner to start to shift their IL form, or at the very least question their IL form. However, tests employed to date would not pick up on these types learning.

The concerns with previous recast studies are neither few nor trivial. The majority of previous studies have unfortunately involved uptake, dyadic interaction and/or intensive recasts directed at pre-selected linguistic target forms. As already mentioned, uptake is hardly a reliable measure of noticing, let alone L2 learning; dyadic interaction provides more individual attention than is seen in natural L2 classrooms; and intensive recasts directed at pre-selected linguistic target forms do not occur naturally in L2 classrooms and may increase attention to those forms beyond what would normally occur in the classroom. To my knowledge, there is only one published study that has examined incidental, extensive recasts in non-dyadic interaction (Loewen & Philp, 2006), and the testing procedure in this study is not without problem. This indicates a clear need for more (pre-)/post-tests recast studies which are not dyadic, do not involve a pre-selected target linguistic form, and do not provide only intensive recasts. After all, if we are not studying recasts the way that they occur naturally in L2 instruction, we will never know

the actual benefit of recasts to L2 learners. In addition, we will not be able to make suitable recommendations regarding recasts to L2 teachers.

2.8 Purpose of the Study

The purpose of the present study is to examine the effectiveness of recasts while addressing some of the gaps and issues from previous studies. Specifically, the present study adds to the (small) body of research on incidental, extensive recasts in non-dyadic interaction. In addition, the present study attempts to overcome some of the problems found with the individualized post-tests used in Loewen and Philp (2006) by employing a new testing methodology: *stimulated correction* (explained in detail in Chapter 3). The *stimulated correction* testing method was designed to have the following advantages over previous measures of the effectiveness of recasts:

- It is a timed, oral test (both stimuli and response); this is designed to put the student in a situation similar to the situation in which the student initially made the error.
- The context of the errors is provided during the testing phase.
- Students view and listen to their errors exactly as they made them during the initial interaction, not spoken by a NS examiner.
- It allows for comparisons between each students' performance on those errors that received recasts and those errors that did not receive recasts.
- It measures students' ability to detect and correct errors separately; this avoids the "all or nothing" view of effectiveness in L2 learning and provides more fine-tuned testing.
- It allows for the measure of reaction times when students are detecting errors in

video clips of their speech.

2.9 Research Questions

Employing a new methodology, *stimulated correction*, along with currently used methodology of individualized written post-tests, the study addressed the following research questions:

1. Are errors that receive recasts detected more often than errors that do not receive recasts?
2. Are errors that receive recasts corrected more often than errors that do not receive recasts?
3. Are there any differences in students' reaction times in detecting errors for episodes that receive recasts and those that do not?
4. Does the type of test make a difference in the results (i.e. stimulated correction vs. written test)?

Chapter 3 – Methodology

3.1 Design

The current study employed an experimental within-subject research design (a design in which each participant served in more than one experimental condition). Small groups of adult ESL students participated in an oral task with a teacher (the researcher) in which some of each student's errors received recasts and some of their errors received no feedback (were ignored). One day after participation in the small-group session, students returned for individualized testing using a stimulated correction task and a written test. In total, seven groups of students completed the study. Each group received identical treatment, and the results of the groups were pooled for analysis. A pilot study with 4 students was conducted two months before the main study to test the procedure. The results from the pilot study were used to make small adjustments to the procedure, as discussed in the various sections below. The data from the pilot study was not included in the analysis and results discussed in chapter four.

3.2 Participants

In total, 26 students participated in the main study. However, three of the students failed to complete day two of the study and so their data were excluded from the analysis. The students were all enrolled in intermediate-level classes (level 330 and 410) at the English Language Centre at the University of Victoria at the time of the study. Intermediate-level students were chosen as their English proficiency is high enough to allow for a fair amount of discussion, but low enough that they were predicted to make

errors during the oral interaction. Of the 23 students who completed the entire study, 14 were female and 9 were male. They had a mean age of 25.2 years and had been living in Canada for an average of 4.2 months at the time of the study. They were all international students and came from eight different countries (Brazil, China, Japan, Korea, Mexico, Russia, Taiwan, and Turkey). There were seven L1s represented; details of the participants' L1s are included in table 3.1 below. Note that none of the groups were composed of a single L1.

Table 3.1. Small-group information.

Group	Level	Number of students		L1*
		Day 1	Day 2	
A	330	4	4	Chinese (2), Spanish (1), Turkish (1)
B		3	3	Russian (1), Spanish (1), Turkish (1)
C		4	2	Korean (3), Japanese (1)
D	410	3	3	Korean (2), Japanese (1)
E		4	4	Korean (2), Chinese (1), Portuguese (1)
F		5	5	Korean (3), Chinese (2)
G		3	2	Korean (2), Chinese (1),
TOTAL:		26	23	Korean (12), Chinese (6), Japanese (2), Spanish (2), Turkish (2), Portuguese (1), Russian (1)

*The number of students in each group with a given L1 is indicated in brackets.

The students were recruited on a voluntary basis. The researcher visited four ESL classrooms (two 330-level and two 410-level) to explain the study and seek volunteers. After the 5-minute recruitment presentation, students who wanted to volunteer signed up for dates and times for the small group interaction and the testing. Students were assigned to one of seven small groups on a first-come, first-serve basis. Thus, the first students to volunteer were placed into group A, then group B was filled, etc. Each group had a

maximum of five slots available for students to sign up for. Based on poor turnout during the pilot study, it was decided that the sign-up sheet for each group should have five sign-up slots in the hopes that of the five who signed up, at least three would show up for the interaction and testing. Thus, while five students signed up for each group, actual turnout ranged from three to five students for the small-group activity (see table 3.1 for number of students in each group). In exchange for their participation, students received a certificate of participation.

The present study employs a small-group, as opposed to classroom or dyadic, design. The use of small-groups in the present study was designed to mimic the way in which a small group would operate within a classroom. As discussed in Chapter 2, questions have been raised as to the naturalness and generalizability of teacher-student dyadic studies. Therefore, I wanted to avoid the individualized attention that students receive in dyadic studies. However, due to time and equipment limitations, I was not in a position to carry out the present research design in a natural L2 classroom (though I feel it would be feasible to carry out a similar study in a classroom situation). Therefore, similar to Han (2002), the present study examines recasts in the context of small-group interaction. It should also be noted that the number of students in each group in the present study (between three and five; see table 1) is comparable to the Korean EFL “classrooms” in Sheen (2006), which had between four and six students. Therefore, although not as ideal as a natural classroom situation, I feel the small-group organization of the present study is a step in the right direction.

In the present study, the researcher took on the role of the teacher. She had two years full-time EFL teaching experience and three years part-time ESL/EFL teaching

experience at the time of the study. As such, she was familiar and comfortable with taking on the role of a teacher when interacting with ESL students. During the recruitment process, the researcher was introduced to the students as having experience as an English teacher.

3.3 Procedure

Data collection for the seven groups took place over eight weeks (one group per week, with one week break over holidays). The data collection took place over two days for each group. This section describes the procedure for one group to complete the study from beginning to end. This procedure was then repeated with each subsequent group until all seven groups had completed the study.

3.3.1 Day One: Treatment

On day one, the small group of students met with the researcher in a small classroom at the University of Victoria. After signing a consent form, the students filled out a background questionnaire, which included questions about the students' age, nationality, L1, and length of English study (see appendix B). After completion of the background questionnaire, the students participated in an oral small-group activity. For the activity, the students and the researcher sat around a table so that they could all see and hear each other comfortably. The interaction was captured using a digital video camera, which was placed approximately five metres away from the table, and a microphone that was connected to the video camera and placed in the centre of the table.

The small-group activity was designed by the researcher. The activity consisted of one task involving two parts: Jigsaw (see Crookes & Gass, 1993; Johnson, 1981) and decision making (see Crookes & Gass, 1993; Doughty & Pica, 1986). The jigsaw part of the task involved the students sharing information from the CVs of (fictional) people applying for a teaching job. The decision-making part of the task involved the students deciding as a group which applicant to hire for the teaching job.

To introduce the activity, the researcher explained that there is an open position for a new ESL teacher at the students' school and that, as a group, the students needed to share information from the CVs in order to come to a final decision on which teacher to hire. The students were told from the outset that this was a fictitious situation and the CVs did not belong to real people. Each student was given the CV of a person applying for the job. Each student had a different CV (see appendix C for example). The students first had five minutes to look over their CVs and ask the researcher any questions they had about the vocabulary on the CVs. The students were then required to orally share, in the third person, the information on their CV with the group. After the information was shared, the students discussed the pros and cons of each person until they came to a final decision on which person to hire (decision-making). During the decision-making part of the task, the students were encouraged to ask questions of each other and consider all of the information presented on the CVs. They were also allowed to take notes to help them remember information about the other applicants. The average time the groups took to complete the task was 38 minutes; task-completion time ranged from 32 to 45 minutes, with those groups containing more students (and thus more CVs to share) taking more

time than the smaller groups. The task was not designed to elicit any specific linguistic forms.

A task with two parts was chosen for several reasons. First, the jigsaw part was chosen to ensure that all the students were required to participate, thus ensuring each student (even those who are shy) would have opportunities to make errors and receive recasts. The decision-making part of the task was chosen because (a) it would allow for a wider range of linguistic forms to be used than in the jigsaw task, and (b) it was assumed that the need for the students to come to a single decision (from multiple possible outcomes) would allow for more interaction and discussion than a simple opinion-exchange task (Crookes & Gass, 1993).

During the activity, the researcher took on the role of the teacher. This consisted of several forms of involvement in the group discussion. Of primary interest was the provision of recasts. It was planned that the researcher would provide recasts following roughly half of each student's erroneous utterances. To accomplish this, the researcher strove to recast every other erroneous utterance for each student; the remaining errors received no feedback (were ignored). However, it should be noted that the ratio of recasts to ignored errors was not 1:1 for every student as some students produced an odd number of erroneous utterances, and sometimes the researcher inadvertently ignored two errors in a row or provided two recasts in a row. In total, there were 141 erroneous utterances that received recasts and 150 erroneous utterances that received no feedback, leading to a ratio of 1.06 to 1.00.

The recasts in the present study were provided extensively. In other words, they were provided to whatever lexical or morpho-syntactic errors occurred throughout the

interaction. Phonological errors were not recast in the present study, as it is difficult to draw the line between erroneous and correct pronunciations with L2 learners; it can always be argued that their utterances are spoken with a foreign accent and thus, strictly speaking, non-native-like (and considered erroneous by some). In the pilot study, one participant indicated that all of his video clips contained errors, but when asked to orally correct them, he often repeated the same words, indicating that he was attempting to correct some perceived phonological errors. Unfortunately, the analysis of such corrections, and whether they were successful or partial successful or unsuccessful, would have been very tedious and taken away from the main focus of the study. Thus, phonological errors were excluded from receiving recasts or being analyzed in the present study. The recasts provided during the interaction had the following characteristics: (a) immediately followed a student's erroneous utterance, (b) repeated all or part of the student's utterance while reformulating the error(s), (c) did not change or add any information to the learners' target-like portion of the utterance (other than pronouns, which were sometimes changed), and (d) employed a rising intonation (see example 3.a below). Recasts varied both in terms of length (number of words) and number of errors reformulated. In terms of length and number of errors corrected, the researcher provided the type of recast that felt most natural at the time of the interaction (unless otherwise noted, all examples in this chapter are from the present study; erroneous utterances are underlined and recasts are bolded).

- 3.a. S1: They should look the your eye.
 T: **Look you in the eye?**
 S1: In the eye.

A rising intonation has been employed in the recasts in the present study because previous studies have found a beneficial role of rising intonation in recasts. Nassaji (2006) found that recasts+prompts (including rising intonation, stress, and or emphatic words) lead to more repair than recasts–prompts. Loewen and Philp (2006) found that recasts with rising intonation lead to more accurate test scores than recasts with declarative intonation. They concluded this may be because the rising intonation encourages the learner to examine whether the teacher’s recast matches what they wanted to say, thus leading to more advanced cognitive comparison (Ellis, 1994). Thus, rising intonations were provided on all recasts in the present study.

In addition to providing recasts, the researcher was involved in the discussion in several ways, namely asking information-gathering questions, giving her opinion when requested by students, and answering student-initiated questions. The purpose of this involvement was to make the student-teacher interaction as natural as possible. In addition, it was thought that if the only interaction by the teacher consisted of recasts, this might heighten the students’ awareness of the recasts and cause them to pay more attention to the recasts than they would in a natural L2 classroom. During these additional types of involvement, care was taken to avoid any involvement that could be interpreted by the students as reacting to their erroneous utterances. Specifically, while information-gathering questions (such as, “Does she have any experience?” or “Who do you think we should hire?”) were asked by the researcher, the researcher did not ask clarification questions. In addition, the researcher avoided asking questions following student uptake of a recast. As for answering student-initiated questions, the majority of which were vocabulary-based questions that arose during the reading of the CVs. While this

involvement of the researcher can be considered form-focused, none of the episodes containing student-initiated questions were subsequently used in the testing phase (on day two, described below).

3.3.2 Creation of Testing Materials

Each student was tested both orally (stimulated correction task) and in written form (written test). The stimulated correction task was employed to test students' ability to detect and correct errors in their own speech. It was designed to overcome some of the concerns with previous post-testing procedures in extensive recast studies (see section 2.8). Individualized written post-tests were administered in order to compare the results of the stimulated correction task with the results of a previously employed method of language assessment (as used in Loewen 2005; Loewen & Philp, 2006), thus increasing the validity of the results. In addition, several studies have examined the benefits from recasts using both written and oral tests and have found that the benefits of recasts are more pronounced on written tests than on oral tests (Ammar & Spada, 2006; Han, 2002; Lyster, 2004). By employing both oral and written tests in the present study, comparisons can be made between the results of the present study and these previous studies.

In order to create the tests, the video from the small-group discussions was first imported onto a laptop (iBook G4) using iMovie 5.0.2. The small-group interaction was then transcribed by the researcher using standard orthography. After the transcription was complete, the researcher reviewed it and identified three types of episodes: *error+recast*, *error-recast*, and *correct*. Error+recast episodes involved exchanges in which a student made an error and then received a recast from the teacher (see example 3.b). In error-

recast episodes, the student produced an erroneous utterance but the teacher did not provide any feedback to the student (see example 3.c). Correct episodes were those in which the student produced a grammatically correct utterance. Since the utterance was initially correct, no feedback was provided by the teacher (see example 3.d; correct utterances are italicized).

Error+recast episode:

3.b. S4: mmm, sometimes is boring!

T: <<laughing>>

S4: Uh uh

T: So why, what makes it boring? What would you rather do?

S4: Because mmm some people already know teacher teach- uh- teacher taught.

T: **Know what the teacher taught?**

S4: yeah, teacher taught.

Error-recast episode:

3.c. S3: First, maybe the school like uh good at teachings, ESL or what first.

T: Mmmm.

S4: Yeah.

S5: I think uh experience is very important to teach.

S3: Yeah.

S5: I was- So Lisa is best teacher because sh- she had TESL certification.

Correct episode:

3.d. T: Is there any extra information we need to know about Lisa?

S2: Uh ++ she speak uh she + she can speak uh Spanish fluently. *And she can speak a little Japanese.*

T: Okay.

S2: And she- she likes drawing, painting, and doing art.

After all of the error+recast, error-recast, and correct episodes were identified, they were reviewed to determine if the episodes were suitable for testing. Specifically, episodes were excluded from being used in the tests if the student utterance in focus in the episode (a) contained only one word, (b) contained more than two false starts, or (c) could not be clearly heard in the video (i.e. several students were speaking at the same time). In total, 402 episodes suitable for testing were identified. Of these, 111 error+recast episodes, 111 error-recast episodes, and 111 correct episodes were randomly selected to serve as testing episodes. These 333 episodes were used to create both the stimulated correction task and the written test (see sections 3.3.2.1 and 3.3.2.2 for more details). The remaining 69 episodes (3 per student) were selected to serve as practice episodes on the stimulated correction task.

3.3.2.1 Creation of stimulated correction task.

After the 333 test episodes and 69 practice episodes were selected, they were edited using iMovie to create the short video clips that would be the stimulus for the stimulated correction task. Each of the episodes was edited to end with a student's utterance (either erroneous or correct), which was the *utterance in focus* (the utterance that the student would be required to make a grammaticality judgment on and orally correct if necessary). The main purpose of editing the episodes was to remove the recasts so that the student would not have any clues in the video as to the correctness of their utterance. As such, it could be concluded that any learning that occurred as a result of the recast would have taken place during the interaction, not from viewing the recast a second time when they watched the video. The cutting out of the recast in each episode

yielded clips that ended with a student's erroneous utterance. In order to make the clips for the error-recast and correct episodes similar in style, these episodes were also edited to end with the student's utterance (see examples 3.e, 3.f, 3.g; the boxes around the utterances indicate which utterances were included in the edited clips). Finally, an auditory cue, a short "beep", was added to the video clips immediately before the student's utterance in focus. In the examples below, the placement of this auditory cue is indicated by the symbol ♪. The purpose of this cue was to direct the student's attention to the utterance they would be required to judge the grammaticality of and orally correct if they thought it was erroneous. In sum, all of the edited clips shared the following characteristics: (a) included several utterances from the teacher and/or students to provide context, (b) did not include any feedback from the teacher on the correctness of the student's utterance, (c) ended with a student's utterance that was the utterance in focus for the clip, and (d) included an auditory cue immediately before the utterance in focus.

3.e. T: Well, what do you guys think?

S1: I think if, if, uh Lisa can speak another language, she understands another student things ♪ so she can teach very well that country students.

T: **Students from that country?**

S1: Yeah.

3.f. S2: Yeah, I think so.

T: You all agree! For the same reasons or different ones?

S1: Uh, I think uh Philip is more academic than Lisa, but he has enough- he te- ♪ he didn't have enough experience to teach to the student.

T: Mm hmm.

S1: So just uh he tea- he taught reading and writing-

3.g. T: Mm hmm. Okay. So whose work experience do you like the best? Do you think?

S4: I think Lisa.

T: Okay, why?

S4: ♪ *Because she has uh lots of experience teaching ESL.*

Because three years she teach English and Lisa also uh it is relative in researching.

T: Mm hmm.

After all of the clips from a group's interaction session were created, they were sorted by student. This was done because each student was tested on only those clips involving their own utterances, for example, S1 was tested on the clips created from examples 3.e. and 3.f. above, while S4 was tested on the clip from example 3.g. The reason students' were only tested on their own utterances is twofold. First, the fact the student initially made the error indicates an incompleteness in some level of their knowledge of that form, and thus serves as a type of pre-test (see Loewen, 2005; Loewen & Philp, 2006). Second, a correct answer provided by S1 after viewing a clip of another student's error does not indicate that S1 learned from the recast provided to the other student; S1 could have know the correct form even before the interaction.

After the clips for a given student had been created, they were inputted into the testing software SuperLab 4.0 to create the stimulated correction task. A separate SuperLab file was created for each student. Each file consisted of three blocks: a demonstration block, a practice block, and a testing block. Each block consisted of several trials (each trial included one video clip). The demonstration block was the same for all the students, and consisted of five trials; the clips for each trial involved the

researcher interacting with two “ESL students” (they were in fact graduate students from the Linguistics Department). The practice block contained three trials; for these trials, each student practiced on clips of his/her own speech. The number of trials in the testing block varied between students. This was because some students were more active than others during the small-group activity, leading to more clips for those students. Each student received an average of 14.5 trials in the testing block (range was from 7 to 21).

The set-up of each trial in the experiment was identical and consisted of four components: (1) a screen with the words “The video will begin in 3 seconds”, which was displayed for three seconds, (2) the playing of a video clip, (3) a screen with the words “Press ERROR or NO ERROR”, which was displayed until the student pressed either the “ERROR” key or the “NO ERROR” key on the keyboard, and optionally (4) a screen with the words “Now correct it”, which was displayed for 8 seconds (the pilot study showed that 8 seconds was an adequate amount of time for the students to orally correct their utterances). This screen would appear only if the student had pressed the “error” key; if the student pressed the “no error” key, screen number 4 would not appear and the trial would end with component 3. At the end of each trial, the next trial would begin immediately. Each trial was presented only once as the pilot study confirmed that students felt comfortable completing the task after viewing the video only once. The trials were presented in random order within each block. Reaction time (in milliseconds) was measured between the end of the video clip and the time at which the student pressed the “error” or the “no error” key.

3.3.2.2 Creation of the written test.

After the interaction had been transcribed and the three types of episodes identified, the researcher created a written test, based on the same episodes as the stimulated correction task, for each student. The researcher took the episodes for each student and typed the utterances in focus (i.e. those occurring after the ♪ in examples 3.e, 3.f, 3.g) from all of his/her episodes into Microsoft Word. If the student made a false start and then changed their utterance, only the final portion was included in the test question (i.e. if the student said, “He like- he likes to paint.” the test question would read: He likes to paint.) Each test question consisted of one sentence/utterance. When all the utterances in focus had been written down, they were placed in random order. To the right of each sentence was placed ___ error ___ no error. Two blank lines were also left after each question to allow students’ space for corrections of their error(s). See Appendix D for a sample written test. The written test and the stimulated correction task had the same number of questions for a given student (i.e. if S1 viewed 15 clips in the stimulated correction task, they answered 15 questions on the written test). However, as both the stimulated correction clips and the written test questions were presented in random order, the order of presentation of the utterances in focus was different on the stimulated correction task and the written test.

3.3.3. The Testing Procedure

On day two, which took place the day after the small-group interaction, each student was tested individually in a quiet room. First, the student re-signed the consent

form from day one. Next, the student completed the stimulated correction task. For the stimulated correction task, the student sat in front of a computer screen and listened to the stimulus through headphones. The students' speech during the stimulated correction task was recorded using a digital voice recorder, which was placed beside the computer.

For the training, the researcher explained the procedure as follows (see full script in Appendix E): (1) wait for the video clip. (2) Listen carefully to your own speech during the video clip, paying special attention to your utterance following the "beep". (3) Decide if you think your speech after the "beep" contained any errors. If you think your speech did contain one or more errors, press the "error" key on the keyboard; if think your speech did not contain any errors, press the "no error" key. Do not worry about pronunciation errors. It is important that you try to answer as quickly as you can. (4) If you think your speech contained error(s), correct them by saying out loud the correct way to say it. (5) Wait for the next video. The students were also explicitly told that some of the video clips did contain errors and some of the video clips did not contain errors. After explaining the procedure, the researcher demonstrated the procedure by completing the demonstration block as the student watched. After answering any questions from the student, the researcher told the student that they would complete some practice trials, which would not be recorded. The student completed the three trials in the practice block while the researcher remained silent. Upon completion of the practice trials, the researcher asked the student if they would like to try the practice trials again or if they were ready to begin the real trials. The students were permitted to repeat the practice trials until they felt comfortable with the procedure.

When the student felt ready, the testing block was administered. During the testing block, the student viewed all of the trials (n ranged from 7 to 21) in random order without breaks. Students took between approximately 5 and 12 minutes to complete the task. No feedback on the correctness of their responses was provided to the students during the task, and the researcher sat at the back of the room (behind the student) to avoid unconsciously providing visual feedback.

After the stimulated correction task had been completed, the student was given the written test and told it contained written versions of the utterances that they had just viewed in the video clips. They were told to read each sentence and decide if it contained any errors. If they thought it did have one or more errors, they were to place a checkmark beside “error”. If they thought it did not contain any errors, they were to place a checkmark beside “no error”. Their second task was to correct any errors by crossing out, adding, or rewriting. The students were given as much time as they needed to complete the written test. It took them between approximately 7 and 20 minutes to complete the test.

Following the completion of the written test, the researcher went over the questions from the written test with the student, providing feedback on their performance. The researcher gave the student the correct answers for the questions that they had answered incorrectly. Finally, the researcher thanked the student for their participation and gave them a certificate of participation.

Chapter 4 – Data Analysis and Results

4.1 Data Analysis

Following the end of the data collection period, the interaction data, which consisted of the student utterances and the teacher-provided recasts, was transcribed and students' responses on the stimulated correction task and the written test were scored. After the coding and scoring was complete, the data was inputted into SPSS (15.0) and descriptive and inferential statistical tests were carried out.

4.1.1 Interaction Data

The data from the interaction (day 1 of the study) consisted of students' initial utterances (those utterances on which the student was required to make a grammaticality judgment on day 2) and teacher-provided recasts. The students' utterances and the teacher's recast (if provided) were transcribed in Microsoft Excel (v.X for Mac.). The purpose of transcribing this data in Excel was to allow for easy comparisons between each utterance, (recast), and student's modifications on the stimulated correction task and the written test.

4.1.2 Scoring the Stimulated Correction Task

The scoring of the stimulated correction task involved analyzing error detection responses, error correction (modification) responses, and reaction times for error detection. The analysis of each of these types of data is described below.

As the students completed the error detection portion of the stimulated correction task, the SuperLab software recorded the key (“error” or “no error”) that each student pressed after viewing each of the video clips (episodes). Since the accuracy (correctness) of the key pressed depended on the type of episode (i.e. pressing the “error” button was an accurate response for error+recast or error-recast episodes, but it was not an accurate response for correct episodes), SuperLab was programmed to also record an “accuracy of response” answer for each time the student pressed a key. Specifically, the response would be coded as correct in the following cases:

- (a) The episode was error+recast and the student pressed “error”
- (b) The episode was error-recast and the student pressed “error”
- (c) The episode was correct and the student pressed “no error”

The response would be coded as incorrect in the following cases:

- (a) The episode was error+recast and the student pressed “no error”
- (b) The episode was error-recast and the student pressed “no error”
- (c) The episode was correct and the student pressed “error”

The second portion of the stimulated correction task required that students orally correct (modify) any utterances they perceived as erroneous during the error detection task. Their modifications were recorded using a digital voice recorder. The audio files of the students’ error modifications were imported into a G4 iBook and were played using QuickTime 7.1.3. They were then transcribed using standard orthography. After being transcribed, each response was coded. The coding involved comparing the student’s utterance from the interaction with the student’s modification of their utterance during the error correction part of the stimulated correction task. Students’ modifications were

coded as *successful modification*, *partially successful modification*, *unsuccessful modification*, or *no modification*. Descriptions and examples of these four types of modification are provided below (in the examples, students' modifications are indicated in bold italic).

Successful modification: The student corrected all the error(s) in their utterance.

4.a. Student's utterance during interaction activity:

Important one I see the person was good prepare this area.

Student's response in stimulated correction task:

It's important for the person to have experience in this area.

4.b. Student's utterance during interaction activity:

She like drawing, painting, and doing art.

Student's response in stimulated correction task:

She *likes*. Not she like.

Partially Successful modification: The student corrected only one/some of the error(s)

or they improved on the utterance but not to a point that it was completely grammatical (i.e. inserted a needed verb, but made a mistake on the conjugation).

4.c. Student's utterance during interaction activity:

She is over 5 years English teaching experience, experience.

Student's response in stimulated correction task:

She *already have* over 5 years teaching experience.

4.d. Student's utterance during interaction activity:

And the teacher have to know another country's people.

Student's response in stimulated correction task:

The teacher *has* to know.

Unsuccessful modification: The student failed to correct the errors or improve on the grammaticality of the utterance in any way.

4.e. Student's utterance during interaction activity:

She was a award.

Student's response in stimulated correction task:

She was award. *[the student deleted 'a']*

4.f. Student's utterance during interaction activity:

And she has a more experience- more work experience teacher.

Student's response in stimulated correction task:

I think she's *a high experience* teacher.

No modification: The student did not make any changes to their utterance. This included instances where the student repeated word-for-word their initial utterance and instances where the student said they did not know how to correct their utterance.

4.g. Student's utterance during interaction activity:

And then she taught ESL to adult in Korea.

Student's response in stimulated correction task:

I don't know how to make a correct sentence.

4.h. Student's utterance during interaction activity:

To learn how teaching people.

Student's response in stimulated correction task:

To learn.

The above coding scheme was applied to all three types of episodes. However, it should be noted that the meaning of the types of modification was slightly different for the correct episodes. Specifically, for correct episodes, the coding of "successful" meant that the student had changed the (already correct) utterance in such a way that it could also be considered grammatically correct (see example 4.i. below). The coding of a modification as *unsuccessful* meant the student changed the correct utterance into an incorrect one. In addition, the coding of a modification as *partially successful* was not applied to correct episodes, as it is not possible to take a correct episode and change it to make it only partially correct; it would simply be ungrammatical (unsuccessful). As such, student modifications to correct episodes were coded into three categories: successful, unsuccessful, or no modification.

4.i. Student's utterance during interaction activity:

Our pronunciation is a little bit weird.

Student's response in stimulated correction task:

Our pronunciation is a little bit *strange*.

Student modification on error+recast episodes were further coded to allow a more detailed analysis of how recasts may have influenced students' ability to modify their errors. Thus, for error+recast episodes only, successful modifications were coded as

either successful/same as recast (meaning the student modified the utterance in the same way as the recast had the previous day; see example 4.j) or successful/different than recast (meaning the student modified the utterance in a different way than the recast had the previous day; see example 4.k). Partially successful modifications were coded as either partial/same as recast (example 4.l) or partial/different than recast (example 4.m).

4.j. Student's utterance during interaction activity:

After that he researched in the linguistic department in 2003
between 2006.

Teacher-provided recast:

So, between 2003 and 2006?

Student's response in stimulated correction task:

Between 2003 and 2006. (successful/same as recast)

4.k. Student's utterance during interaction activity:

In Philippine is more few people.

Teacher-provided recast:

Fewer people?

Student's response in stimulated correction task:

In the Philippines, the class has few people.

(successful/different than recast)

4.l. Student's utterance during interaction activity:

He have one more year teach English.

Teacher-provided recast:

He has taught English for one year?

Student's response in stimulated correction task:

He has, he has. (partial/same as recast)

4.m. Student's utterance during interaction activity:

He had the research assistant.

Teacher-provided recast:

He was?

Student's response in stimulated correction task:

He *works a* research assistant. (partial/different than recast)

Table 4.1 summarizes the similarities and differences in the coding schemes for the three types of episodes. For some of the statistical analyses presented in the results section in this chapter (i.e. comparing results for error+recast and error–recast episodes), successful/same as recast was combined with successful/different than recast (and labeled *successful*) and partial/same as recast was combined with partial/different than recast (and labeled *partial*).

Table 4.1. Coding of nature of modification according to episode type.

episode	<i>successful</i>		<i>Partially successful</i>		<i>unsuccessful</i>	<i>no modification</i>
	recast	other	recast	other	yes	yes
error+recast	recast	other	recast	other	yes	yes
error-recast	yes		yes		yes	yes
correct	yes		no		yes	yes

4.1.3 Scoring of Written Test

The scoring of the test involved analyzing error detection responses and error correction (modification) responses. The analysis of each of these types of data is described below.

Error detection responses were in the form of checkmarks beside either “error” or “no error” on the written tests. The researcher inputted this data into Excel in a similar manner to the coding of the error detection responses from the stimulated correction task: for error+recast and error-recast episodes, a correct response was a checkmark beside “error”. For correct episodes, a correct response was a checkmark beside “no error”

Students’ (written) corrections (modifications) of their utterances on the written tests were inputted into the Excel file using standard orthography. The responses on the written test were then compared to the student utterances from the interaction and were coded using the same scheme as for the stimulated correction task: successful modification, partially successful modification, unsuccessful modification, or no modification. The fine-tuned coding for to the error+recast episodes (same as recast/different than recast) was also applied.

4.1.4 Reaction Times

Finally, reaction times for the error detection task were measured and recorded by Superlab for each episode. The reaction time was the time from the end of the video clip until the moment when the student pressed the “error” or the “no error” key. Superlab recorded these reaction times in milliseconds, but they were converted to seconds for easy reading and analysis.

4.2 Error Detection Results

4.2.1 Stimulated Correction Task

Student responses on the error detection component of the stimulated correction task are presented as cross tabulated frequencies in table 4.2 (see also figure 4.1). Please note that for error+recast (hereafter referred to as +recast) and error-recast (hereafter referred to as -recast) episodes, *errors detected* indicates correct responses while *errors not detected* indicates incorrect responses. For correct episodes, *errors detected* indicates incorrect responses and *errors not detected* indicates correct responses. The results showed that students responded significantly differently to the three types of episodes [$\chi^2(2, N = 333) = 15.10, p = .001$]. Specifically, students responded “error” more often for +recast and -recast episodes than for correct episodes, which was expected.

Table 4.2. Students’ responses on the error detection portion of the stimulated correction task.

	errors detected		errors not detected	
	<i>n</i>	%	<i>n</i>	%
+ recast	85	76.6	26	23.4
- recast	79	71.2	32	28.8
correct	53	47.7	58	52.3
total	217	65.2	116	34.8

$\chi^2(2, N = 333) = 15.10, p = .001$

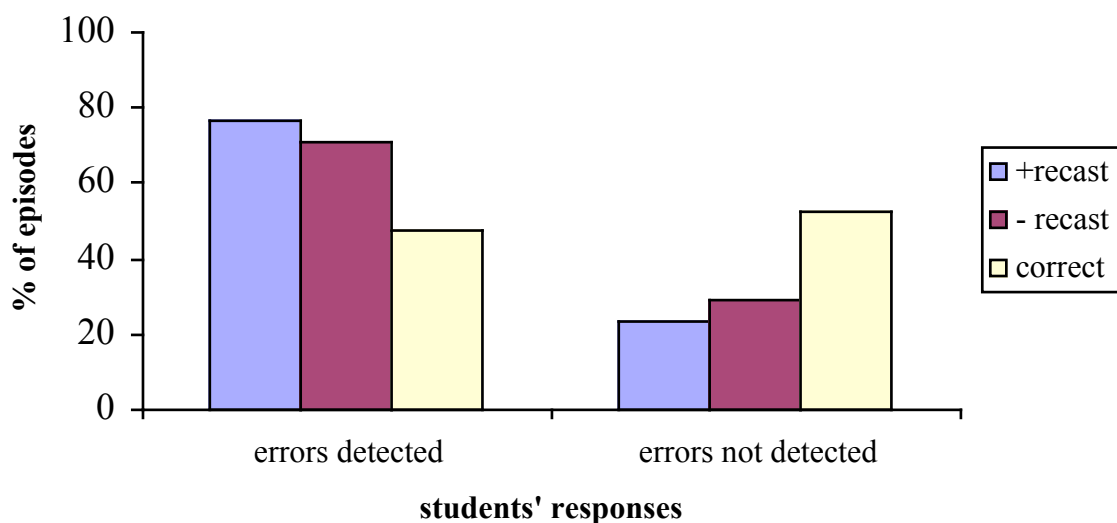


Figure 4.1. Students' responses on the error detection portion of the stimulated correction task.

For the episodes that did contain errors, the +recast and –recast episodes, students detected well over two thirds of the errors (72.9%, $n = 164$). Students detected a higher percentage of errors in the +recast episodes (76.6%) than in the –recast episodes (71.2%). This suggests that the recasts were of benefit to students, assisting them in detecting errors in their own speech. However, chi-square analysis showed that the difference in error detection rates between +recast and –recast episodes did not reach statistical significance [$\chi^2 (1, N = 222) = .840, p = .359$].

For the episodes that did not contain errors (i.e. correct episodes), students correctly responded that the episodes contained “no error” 52.3% of the time. However, they incorrectly identified 47.7% of the episodes as containing errors. In other words, just under half of the correct episodes were interpreted by students as being erroneous even though the episodes in fact contained no errors.

Thus, in response to the question of whether errors that received recasts would be detected by students more often than errors that did not receive recasts, the results of the stimulated correction task showed that errors that received recasts were detected slightly more often than errors that did not receive recasts. However, the difference was not statistically significant.

4.2.2 Written Test

Student responses on the error detection component of the written test are presented in table 4.3 and figure 4.2. As with table 4.2, for +recast –recast episodes, *errors detected* indicates correct responses while for correct episodes, *errors detected* indicates incorrect responses. The results showed that students responded significantly differently to the three types of episodes [$\chi^2 (2, N = 333) = 46.636, p = .000$]. Similar to the stimulated correction task, students responded “error” more often for +recast and –recast episodes than for correct episodes on the written test.

Table 4.3. Students’ responses on the error detection portion of the written test.

	errors detected		errors not detected	
	<i>n</i>	%	<i>n</i>	%
+ recast	93	83.8	18	16.2
- recast	91	82.0	20	18.0
correct	59	53.2	52	46.8
total	184	82.9	38	17.1

$\chi^2 (2, N = 333) = 46.636, p = .000$

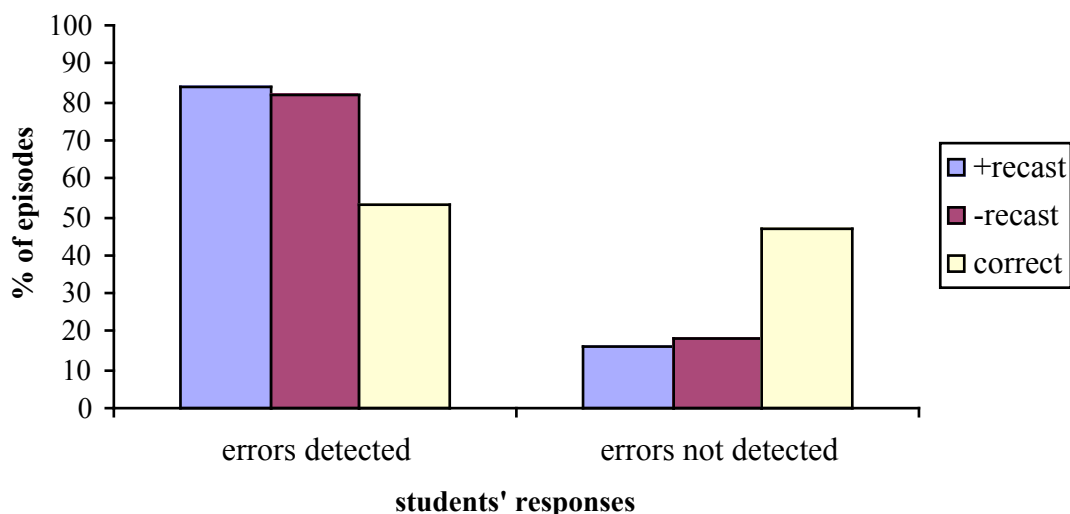


Figure 4.2. Students' responses on the error detection portion of the written test.

For the episodes that did contain errors, the +recast and –recast episodes, overall students detected a high percentage of the errors (82.9%, $n = 184$). They detected slightly more errors in +recast episodes (83.8%) than in –recast episodes (82.0%), but when a chi-square was done, it was found that this difference was not significant [$\chi^2(1, N = 222) = .127, p = .722$].

The written test error detection results for correct episodes showed that students correctly perceived the episodes as containing no errors only 46.8% of the time. The rest of the time (53.2%), they (incorrectly) indicated that the episodes contained error(s).

Thus, in response to the question of whether errors that received recasts would be detected by students more often than errors that did not receive recasts, the results of the written test show that errors that received recasts were detected slightly more often than errors that did not receive recasts. However, the difference was not statistically significant.

4.2.3 Comparison Between Stimulated Correction Task and Written Test

The error detection results from the stimulated correction task and the written test are compared in table 4.4 and figure 4.3. It can be seen that students detected more errors on the written test than on the stimulated correction task for +recast (+7.2%), –recast (+10.8%), and correct episodes (+5.5%). It is not surprising that the students detected more errors in their +recast and –recast episodes on the written test. It can be argued that this increase in accuracy may be because they had more time to consider their utterances on the written test. Another possible explanation is that since the written test was administered after the stimulated correction task, we would expect there to be some practice effect. However, when the results of the correct episodes are also taken into account, it becomes obvious that there may be another factor at play. Surprisingly, on the correct episodes we did not see improved performance from the stimulated correction task to the written test. Instead, students increased their number of (incorrect) “error” responses for correct episodes from the stimulated correction task to the written test (+5.5%). Since students increased their number of “error” responses on the written test for both episodes that did contain errors and those that did not, we cannot assume that students’ apparent improved performance for +recast and –recast episodes on the written test actually indicates an improvement in students’ ability to detect their errors. Instead, we must consider the possibility that (at least some of) the increase in the number of detected errors was due to a bias to answer “error” more often on the written test than on the stimulated correction task.

Table 4.4. Gains/losses in number of “error” responses from the stimulated correction task to the written test.

	stimulated correction		written test		difference	
	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%
+ recast	85	76.6	93	83.8	+8	+7.2
- recast	79	71.2	91	82.0	+12	+10.8
correct	53	47.7	59	53.2	+6	+5.5
total	217	65.2	243	73.0	+26	+7.8

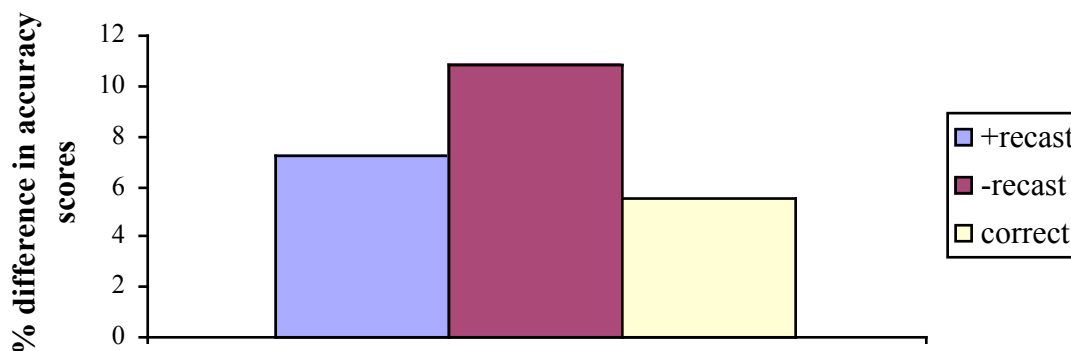


Figure 4.3. Gains/losses in number of “error” responses from the stimulated correction task to the written test.

It should also be pointed out that the differences between stimulated correction scores and written test scores were not equal across all episode types. Namely, as table 4.4 and figure 4.3 show, for +recast episodes students were 7.2% more accurate on the written test than on the stimulated correction task, for –recast episodes students were 10.8% more accurate on the written test, and for correct episodes students were 5.5% less accurate on the written test. There are several possible explanations for this difference between +recast and –recast episodes. First, the results may indicate that the benefits from recasts on students’ ability to detect their errors may be more apparent on an oral test than on a written test. A second possibility is that students were already confident in

their answers for +recast episodes, and thus were not affected by the bias to respond “error” as much on +recast episodes as they were on –recast episodes.

While overall differences in error detection accuracy rates between the stimulated correction task and the written test have shown some variation between the episode types, examining in which ways students changed their responses from one test to the other presents more detailed information. Table 4.5 shows how students’ responses changed or remained the same from the stimulated correction task to the written test. Of the episodes for which students correctly responded “error” on the stimulated correction task, they (incorrectly) changed their responses slightly more often for –recast (12.7%) than +recast episodes (11.8%). The reverse pattern was found when students (correctly) changed their answer from “error” to “no error”, with students changing their mind slightly more often for +recast episodes (69.2%) than for –recast episodes (68.8%). These results indicate that for +recast episodes (as compared to –recast episodes), students were less likely to change their correct responses, and more likely to change their incorrect responses.

As for correct episodes, students frequently changed their response. When students correctly responded “no error” on the stimulated correction task, they (incorrectly) changed their answer to “error” on the written test 42.3% of the time. When they were incorrect on the stimulated correction task (“error” response”), they (correctly) changed their response to “no error” only 35.8% of the time. In other words, they changed more of their correct answers to incorrect ones than vice versa. This suggests that the students were likely not very confident in their responses on correct episodes, regardless of whether they initially thought they contained error(s) or not.

Table 4.5. Changes in error detection responses from the stimulated correction task to the written test.

		+recast		-recast		correct		total	
		<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%
“error” on stimulated correction task	“error” on written test	75	88.2	69	87.3	34	64.2	178	82.0
	“no error” on written test	10	11.8	10	12.7	19	35.8	39	18.0
	total	85	100	79	100	53	100	217	100
“no error” on stimulated correction task	“error” on written test	18	69.2	22	68.8	25	42.3	65	55.6
	“no error” on written test	8	30.8	10	31.2	34	57.7	52	44.4
	total	26	100	32	100	59	100	117	100

Thus, in response to the question of whether there are any differences between the results of the stimulated correction task and the written test, it was found that students did respond differently on the two tasks. Students produced more “error” responses on the written test, regardless of episode type. In addition, it was found that students were less likely to change their correct responses and more likely to change their incorrect responses for +recast episodes than for –recast episodes. Results also showed that students frequently changed their responses on correct episodes from the stimulated correction task to the written test.

4.2.4 Summary

The results from the error detection task demonstrate that students indicated there were errors in +recast and –recast episodes significantly more often than in correct episodes on both the stimulated correction task and the written test. In +recast and –recast

episodes, they were able to detect over two thirds of the errors on both the stimulated correction task and the written test. There was a trend for students to detect more errors in +recast episodes than in –recast episodes, but this difference was not significant on either the stimulated correction task or the written test. A comparison between the results of the two tests show several things: (a) Students were more accurate on the written test, but this was due at least in part to a bias to answer “error” on the written test, (b) there was a greater difference between the responses in +recast and –recast episodes on the stimulated correction task than on the written test, and (c) students were slightly more likely to correctly change and less likely to incorrectly change their responses for +recast episodes than for –recast episodes. Finally, the results for correct episodes showed that students often thought that these episodes contained errors, and that students performed less accurately on the written test than on the stimulated correction task. Results also showed that students often changed their responses on correct episodes from the stimulated correction task to the written test, suggesting that they were not very confident in their responses.

4.3 Error Correction Results

4.3.1 Stimulated Correction Task

Table 4.6 shows the cross-tabulated frequencies for error modification patterns on +recast, -recast, and correct episodes on the stimulated correction task. A chi-square analysis of the modification patterns between the three types of episodes found significant differences [χ^2 (6, N = 217) = 18.73, $p = .005$] among them, which indicates

that the students responded differently when modifying utterances that did indeed contain errors versus utterances that did not contain errors. While the absence of partial modifications for correct episodes is likely partially responsible for this significant difference, the correct episodes also had notably more successful modifications following correct episodes (39.6%) than +recast and –recast episodes (25.9% and 17.7% respectively).

Table 4.6. Error modification patterns (stimulated correction task).

	<i>total</i>	<i>successful</i>		<i>partial</i>		<i>unsuccessful</i>		<i>not modified</i>	
		<i>n</i>	<i>%</i>	<i>n</i>	<i>%</i>	<i>n</i>	<i>%</i>	<i>n</i>	<i>%</i>
+ recast	85	22	25.9	19	22.4	32	37.6	12	14.1
- recast	79	14	17.7	16	20.3	32	40.5	17	21.5
correct	53	21	39.6	NA	NA	21	39.6	11	20.8
total	217	57	26.3	35	16.1	85	39.2	40	18.4

$\chi^2 (6, N = 217) = 18.73, p = .005$

Note. Percentages represent percentage of those episodes in which errors were detected.

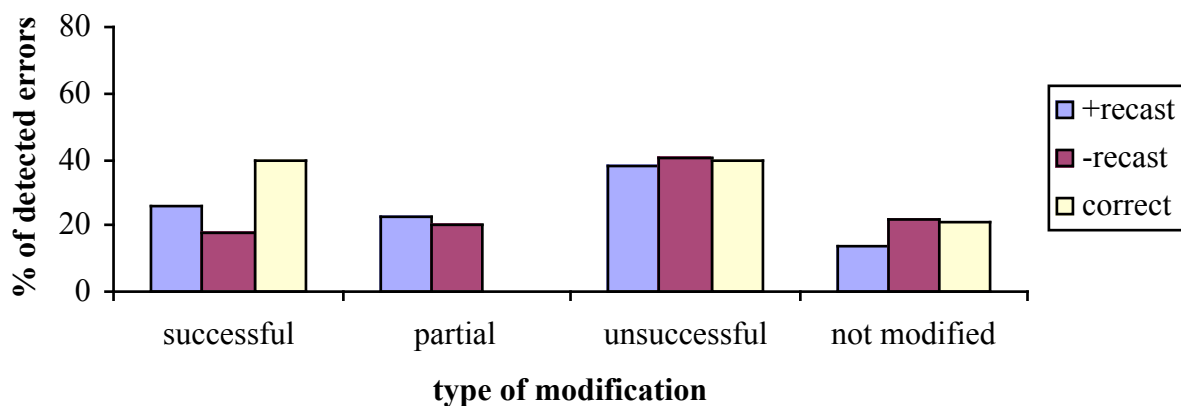


Figure 4.4. Error modification patterns on stimulated correction task.

It was also found that students successfully modified a greater percentage of the errors in the +recast episodes (25.9%) than in the –recast episodes (17.7%). The same trend was found for partially successful modification; errors from +recast episodes were

(slightly) more likely to be partially modified (22.4%) than those from –recast episodes (20.3%). As for unsuccessful modifications, it was found that +recast episodes resulted in fewer unsuccessful modifications (37.6%) than -recast episodes (40.5%). Finally, it was found that some errors were not modified by students even though they identified the episodes as containing error(s). This rate was higher for –recast episodes (21.5%) than for +recast episodes (14.1%). Although the differences in modification patterns in table 4.6 and figure 4.4 show a benefit from recasts, namely that students performed more favorably (i.e. more successful and partially successful modifications, fewer unsuccessful modifications and absences of modification) on +recast episodes than on –recast episodes, a chi-square analysis found that this difference between +recast and –recast episodes was not significant [$\chi^2 (3, N = 164) = 2.681, p = .443$].

Modification patterns for correct episodes are also presented in table 4.6 and figure 4.4. It should be noted that, as explained in chapter 3, modifications of correct episodes were coded as (also) successful, unsuccessful, or no modification; it was not possible to change a correct utterance into a partially correct utterance. It was found that when students (incorrectly) indicated that a correct episode contained error(s), 39.6% of the time they modified that utterance in such a way that yielded an equally acceptable utterance, indicating that they may not be aware that both forms are indeed acceptable. Another 39.6% of the time, the students modified the originally correct utterance in such a way that it became incorrect, indicating that the students may not have been confident in the grammaticality of their original utterance even though it was in fact grammatical. In addition, 20.8% of correct episodes that were identified as containing error(s) were left

unmodified, often with the student still insisting that there was an error but that they did not know how to correct it.

The modification patterns of +recast and –recast episodes were then examined by combining successful with partially successful and unsuccessful with not modified (see table 4.7 and figure 4.5). When this is done, the difference between +recast and –recast episodes became more apparent, with +recast episodes leading to over 10% more successful and partially successful modification (48.2%) than –recast episodes (38.0%); Conversely, +recast episodes lead to over 10% fewer unsuccessful and not modified instances (51.8%) than –recast episodes (62.0%). However, when a chi-square was performed, it showed that this difference still did not reach significance [χ^2 (1, N = 164) = 1.756, $p = .185$].

Table 4.7. Combined error modification patterns, +recast and –recast episodes (stimulated correction task).

	successful & partial		unsuccessful & not modified	
	<i>n</i>	%	<i>n</i>	%
+ recast	41	48.2	44	51.8
- recast	30	38.0	49	62.0
total	71	43.3	93	56.7

$p = ns$

Note. Percentages represent percentage of episodes in which errors were detected.

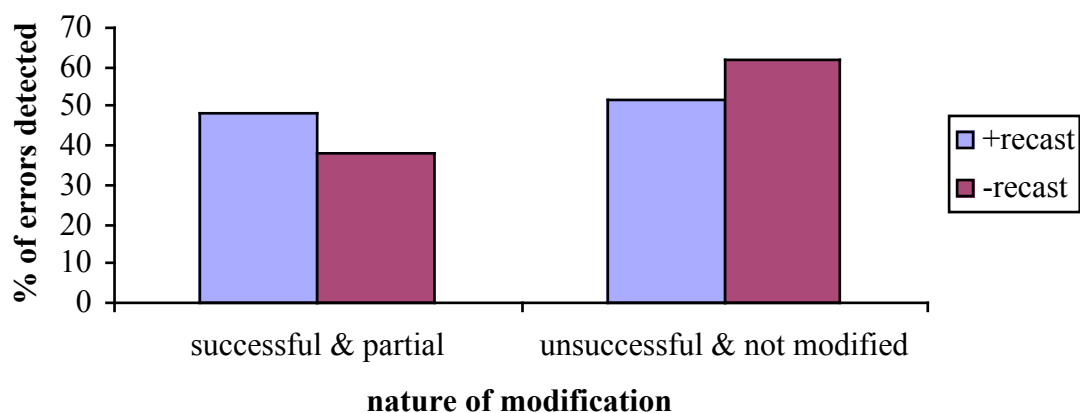


Figure 4.5. Combined error modification patterns for +recast and –recast episodes on the stimulated correction task.

Students' successful and partially successful modifications for +recast episodes were then examined in more detail. Modifications were examined to see if the utterances were modified in the same manner as the recast had been provided or if they were modified in some different way. The results presented in table 4.8 and figure 4.6 show that over two thirds of the time when students successfully or partially successfully modified their utterance, they did so in the same manner as the recast had been provided the previous day (68.3%). Examining only the partially successful modifications, it was found that modifications were split fairly evenly between those that were the same as the recast and those that were different (52.6% and 47.4% respectively). However, the results from the successful modifications were quite different; 81.9% of the successful modifications were done in the same manner as the recast and only 18.1% of the successful modifications were done in a way different than the recast. A chi-square was performed and it was found that the difference between same as recast modifications and different than recast modifications was significant ($\chi^2 [1, N = 41] = 4.011, p = .045$), indicating that when students successfully modified their utterances in +recast episodes,

they were significantly more likely to do so in the same way as the recast had been provided.

Table 4.8. Same as recast vs. different than recast successful and partially successful modifications on +recast episodes (stimulated correction task).

	same as recast		different than recast	
	<i>n</i>	%	<i>n</i>	%
successful	18	81.9	4	18.1
partially successful	10	52.6	9	47.4
total	28	68.3	13	31.7

$\chi^2 (1, N = 41) = 4.011, p = .045$

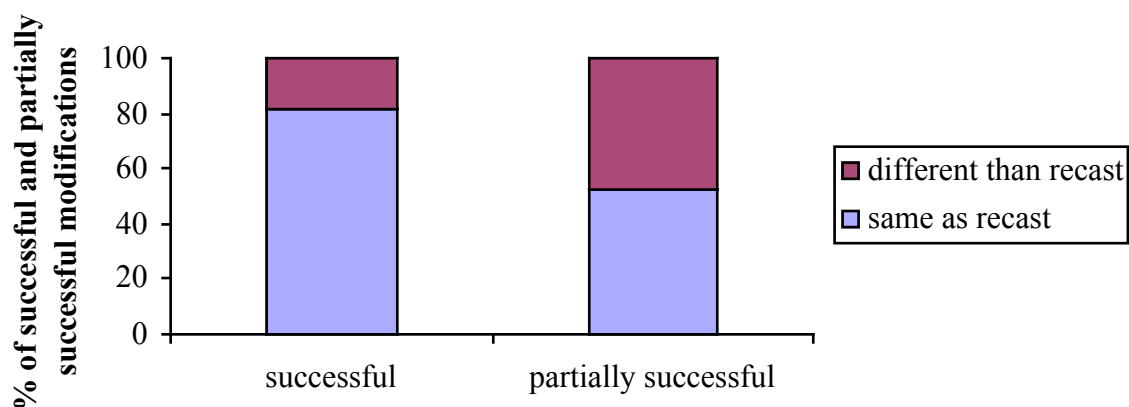


Figure 4.6. Same as recast vs. different than recast successful and partially successful modifications on +recast episodes (stimulated correction task).

The fact that students were significantly more likely to successfully and partially successfully modify their utterances in the same way as the recast is significant given that many of students' erroneous utterances could have potentially been successfully modified in several different ways. Example 4.m. on the next page shows the erroneous utterance (underlined) and the recast (bolded) from the interaction, along with the student's modifications on the stimulated correction task. The example also shows that the modification of changing the verb to *was* is not the only possible way to successfully

modify the erroneous utterance; (at least) four additional successful modifications are possible. The fact that the student chose to modify their error in exactly the same manner as the recast had been provided may suggest that the student remembered the correction provided in the recast, and then used that information on the stimulated correction task. This would mean that the student was able to benefit from the correct form (positive evidence) provided in the recast, and still remember that correct form one day later.

4.m. *Interaction:*

S: Yeah, from 2003 to 2006 she has radio show host.

T: **She was a radio show host?**

Stimulated correction modification:

S: She was a radio show host..

Other possible modifications:

- i. From 2003 to 2006 she had a radio show.
- ii. From 2003 to 2006 she hosted a radio show.
- iii. From 2003 to 2006 she had the position/job of radio show host.
- iv. From 2003 to 2006 she was the host of a radio show.

This provides evidence that the recast may have aided the students in their successful modifications by providing them with a correct modification. Had the recast not assisted them in their successful modifications, we would have expected to see similar rates for the same as recast and different than recast modifications, just as we saw for the partially successful modifications.

Thus, in responses to the question of whether errors that received recasts would be accurately corrected by students more often than errors that did not receive recasts, the results of the stimulated correction task showed a trend for students to successfully and partially successfully modify more errors from +recast episodes than from –recast episodes. However, this difference was not statistically significant. Results also showed that when students successfully modified their utterances from +recast episodes, they were significantly more likely to modify them in the same way as the recast had been provided the previous day than in a different way than the recast had been provided.

4.3.2 Written Test

The error modification results for the written test are presented in table 4.9 and figure 4.7. When the results of the three types of episodes were compared, a chi-square found a significant difference [$\chi^2 (6, N = 333) = 38.415, p = .000$]. The most noticeable difference between the modification patterns is found in successful modifications, with correct episodes resulting in significantly more successful modifications than +recast and –recast episodes.

Table 4.9. Error modification patterns (written test).

	<i>total</i>	<i>successful</i>		<i>partial</i>		<i>unsuccessful</i>		<i>not modified</i>	
		<i>n</i>	<i>%</i>	<i>n</i>	<i>%</i>	<i>n</i>	<i>%</i>	<i>n</i>	<i>%</i>
+ recast	93	34	36.6	33	35.5	23	24.7	3	3.3
- recast	91	21	23.1	29	31.9	38	41.8	3	3.2
correct	59	36	61.0	NA	NA	22	37.3	1	1.7
total	243	91	37.4	62	25.5	83	34.2	7	2.9

$\chi^2 (6, N = 333) = 38.415, p = .000$

Note. Percentages represent percentage of those episodes in which errors were detected.

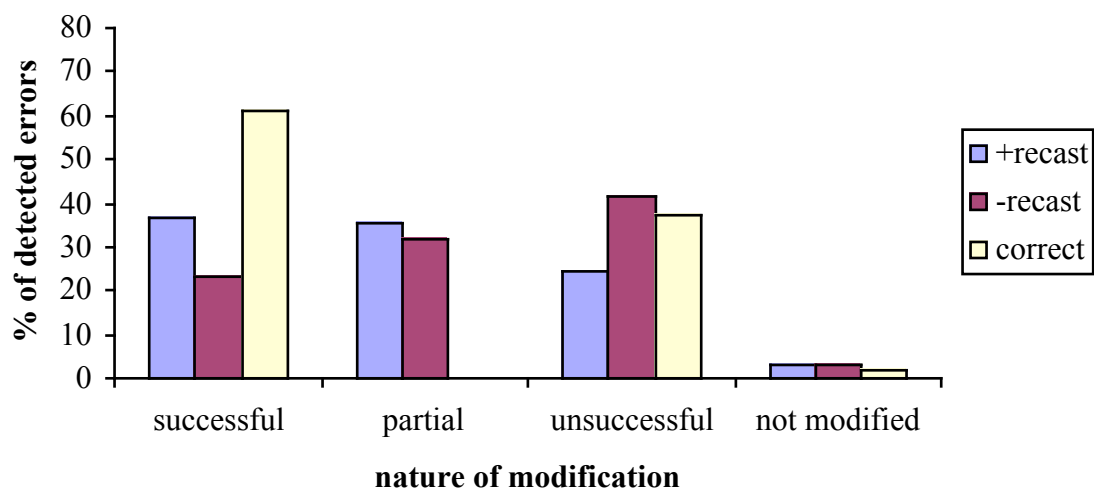


Figure 4.7 Error modification patterns on written test.

As for the comparison between the results of the +recast and –recast episodes, it was found that students successfully modified more of their errors from +recast episodes (36.6%) than from –recast episodes (23.1%). They also partially successfully modified more errors from +recast episodes (35.5%) than from –recast episodes (31.9%). As for unsuccessful modifications, +recast episodes led to far fewer (24.7%) than -recast episodes (41.8%). The percent of detected errors left unmodified was similar for +recast and –recast episodes (3.3% and 3.2% respectively). Thus, the error modification results for +recast and –recast episodes show that students performed better on the +recast episodes than on the –recast episodes, but when a chi-square was performed to see if the differences in modification patterns between +recast and –recast episodes was significant, it was found that they were not [$\chi^2(3, N = 222) = 6.998, p = .072$].

As with the results from the stimulated correction task, the modification patterns of +recast and –recast episodes on the written test were examined in more detail. Table 4.10 and figure 4.8 present the results of combining successful with partially successful

modifications and unsuccessful with not modified responses. As can be seen, +recast episodes resulted in many more successful and partially successful modifications than –recast episodes (72.0% vs. 54.9%), and this difference was found to be significant [$\chi^2(1, N = 222) = 5.807, p = .016$].

Table 4.10. Combined error modification patterns, +recast and –recast episodes (written test).

	successful & partial		unsuccessful & not modified	
	<i>n</i>	%	<i>n</i>	%
+ recast	67	72.0	26	28.0
- recast	50	54.9	41	45.1
total	117	63.6	67	36.4

$\chi^2(1, N = 222) = 5.807, p = .016$

Note. Percentages represent percentage of episodes in which errors were detected.

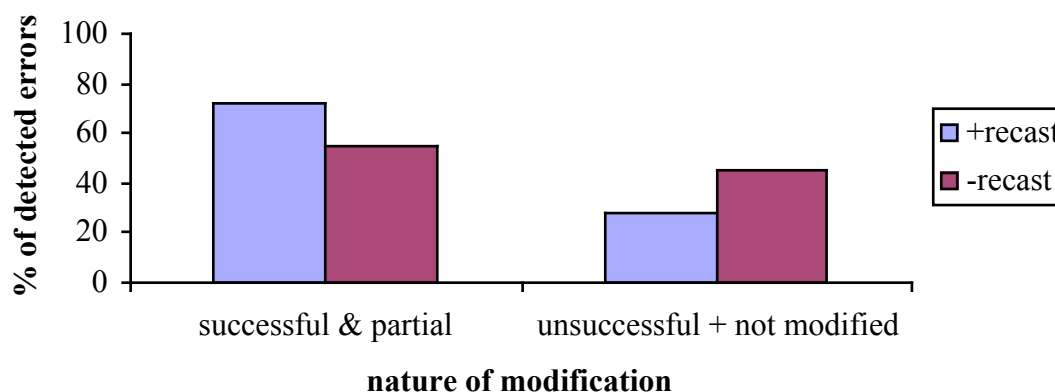


Figure 4.8. Combined error modification patterns for +recast and –recast episodes on the written test.

Student modifications on +recast episodes were also examined in greater detail. Table 4.11 and figure 4.9 display the frequency of whether students' successful and partially successful modifications were done in the same way as the recast had been provided or in a different way than the recast had been provided. It was found that for

both successful and partially successful modifications, students were more likely to modify their utterance in the same way as the recast had been provided (82.4% of successful modifications and 66.7% of partially successful modifications). However, a chi-square showed that these differences did not reach significance [$\chi^2 (1, N = 67) = 2.176, p = .140$].

Table 4.11. Same as recast vs. different than recast successful and partially successful modifications on +recast episodes (written test).

	same as recast		different than recast	
	<i>n</i>	%	<i>n</i>	%
successful	28	82.4	6	17.6
partially successful	22	66.7	11	33.3
total	50	74.6	17	25.4

$\chi^2 (1, N = 67) = 2.176, p = .140$

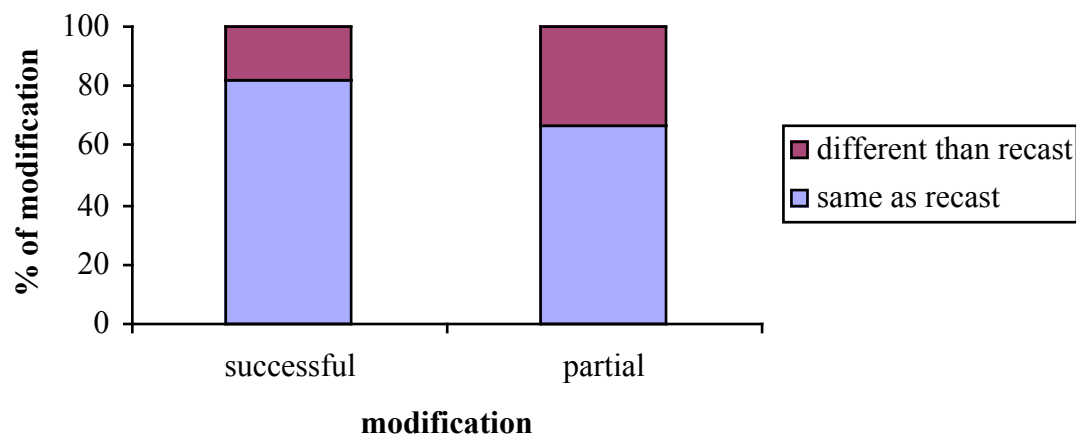


Figure 4.9 Same as recast vs. different than recast successful and partially successful modifications on +recast episodes (written test).

Thus, in responses to the question of whether errors that received recasts would be accurately corrected by students more often than errors that did not receive recasts, the results of the written test showed that students' modifications of their errors on +recast

and –recast episodes were significantly different. Specifically, when successful was combined with partially successful and unsuccessful was combined with not modified, it was found that +recast episodes lead to significantly more successful and partially successful modifications than –recast episodes.

4.3.3 Comparison Between Stimulated Correction and Written Test

Table 4.12 and Figure 4.10 show the gains and losses in error modification from the stimulated correction task to the written test. Positive numbers indicate the students had more of that type of modification on the written test than on the stimulated correction task. Negative numbers indicate that the student had less of that type of modification on the written test than on the stimulated correction task.

Table 4.12. Gains/losses in modifications from stimulated correction task to written test.

	<i>successful</i>		<i>partial</i>		<i>unsuccessful</i>		<i>not modified</i>	
	<i>n</i>	<i>%</i>	<i>n</i>	<i>%</i>	<i>n</i>	<i>%</i>	<i>n</i>	<i>%</i>
+ recast	+12	+10.7	+14	+13.1	-9	-12.9	-9	-10.8
- recast	+7	+5.4	+13	+11.6	+6	+1.3	-14	-18.3
correct	+15	+21.4	NA	NA	+1	-2.3	-10	-19.1

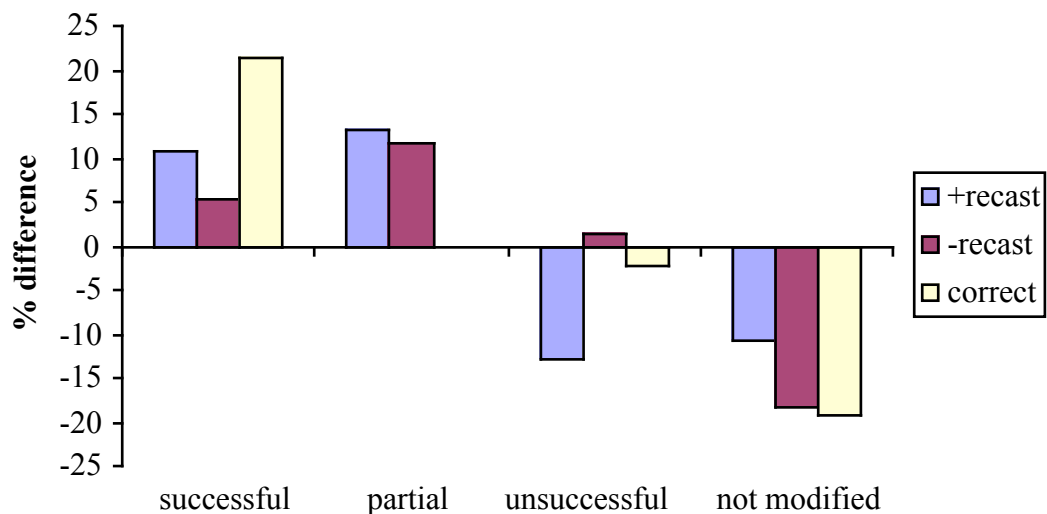


Figure 4.10. Gains/losses in modifications from stimulated correction task to written test.

In general, we see a trend for students to perform better (i.e., positive numbers for successful and partially successful modifications and negative numbers for unsuccessful and not modified responses) on the written test than on the stimulated correction task. However, the effect was not even across the three episode types. For successful modification, while +recast episodes got 10.7% higher scores on the written test than on the stimulated correction task, -recast episodes only gained 5.4%. Correct episodes had the greatest gain, with 21.4%. For partially successful modifications, +recast episodes had slightly higher gains (+13.1%) than -recast episodes (+11.6%). The largest surprise came with unsuccessful modifications. While +recast episodes had 12.9% fewer unsuccessful answers on the written test than on the stimulated correction task, for -recast and correct episodes the number of unsuccessful modifications actually increased on the written test (+1.3% and +2.3% respectively). All three types of episodes had fewer

instances of no modification on the written test, with +recast episodes decreasing by 10.8%, -recast episodes by 18.3%, and correct episodes by 19.1%.

These results show that the three episode types were not affected equally in terms of difference in error modification from the stimulated correction task to the written test. They show that +recast episodes resulted in more improvements from the stimulated correction task to the written test than the -recast episodes did. It is possible that the information that the recasts provided the previous day was indeed stored in the students' memory, but the 8 seconds they were given in the stimulated correction task was not enough time for them to effectively access this stored information. On the written test, however, they did have enough time to access this stored information. Thus, the students would have this resource available to them when they modified their utterances from the +recast episodes on the written test, leading to more successful and partially successful modifications, and fewer unsuccessful modifications. They would not, however, have this resource for the -recast episodes as no recasts had been provided, leading to fewer improvements for the -recast episodes. Another possibility is that when students are corrected using recasts, students may better be able to retrieve this information in written mode than in oral mode (even though the recast was provided orally), leading to more improvement on the written test for +recast episodes. It is also possible that the student had more time to try out various hypotheses on the written test, and when they tried out the one provided by the recast, they immediately knew that it "felt right".

Altogether, not only did students increase their performance from the stimulated correction task to the written test more for +recast episodes than for -recast episodes, but there is also evidence that it was in fact the information provided in the recasts that

helped them do this. Table 4.13 and figure 4.11 show that as students increased their number of successful and partially successful modifications on the written test, the majority of these were produced in the same way as the recast had been provided. Specifically, of the 12 additional successful modifications they produced on the written test, 10 (83.3%) were produced in the same way as the recast but only 2 (16.7%) were produced differently than the recast. For partially successful modifications, of the 14 additional modifications on the written test, 12 (85.7%) were produced in the same way as the recast, but only 2 (14.3%) were produced differently. This suggests that the majority (but not all) of the increase in successful and partially successful modifications on the written test was likely influenced by the information provided in the recasts the previous day. Of course, we must not forget that –recast episodes also had some gains in successful and partially successful modifications (though not as large as +recast episodes), suggesting that a portion of these modifications in +recast episodes may have been identical to the recasts simply by chance.

Table 4.13. Gains in successful and partial successful modifications for +recast episodes from the stimulated correction task to the written test.

	total gains in modifications on written test		same as recast		different than recast	
	<i>n</i>		<i>n</i>	%	<i>n</i>	%
successful	12		10	83.3	2	16.7
partially successful	14		12	85.7	2	14.3
total	26		22	84.6	4	15.4

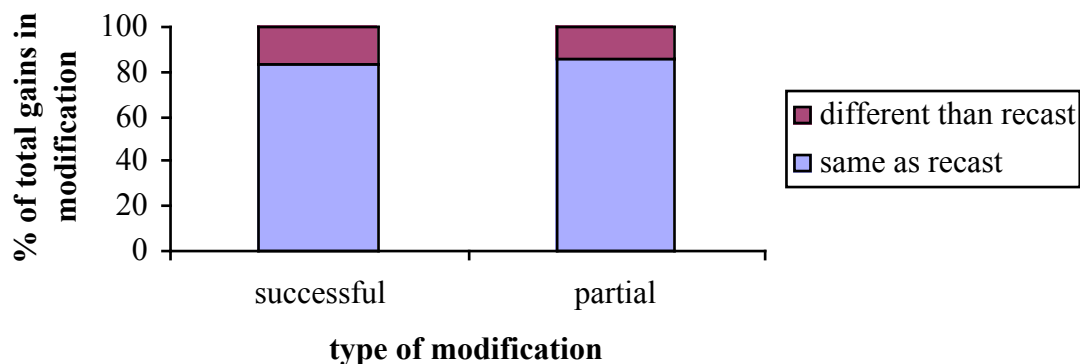


Figure 4.11. Gains in successful and partial successful modifications for +recast episodes from the stimulated correction task to the written test.

Thus, in answer to the research question of whether there are any differences between the results of the stimulated correction task and the written test, it was found that in general, students produced better modifications on the written test than the stimulated correction task and that +recast episodes lead to greater gains than –recast episodes. In particular, +recast episodes had fewer unsuccessful responses on the written test, while –recast episodes had more unsuccessful responses on the written test.

4.3.4 Summary

For the error correction task, students were able to successfully and partially successfully modify a greater portion of their errors on the written test than on the stimulated correction task. On both the stimulated correction task and the written test there was a trend for students to perform better (more successful and partially successful modifications and fewer unsuccessful and not modified responses) on the +recast

episodes than on the –recast episodes, but the difference was not significant. However, there were a couple of significant findings that point towards a benefit from recasts. Namely, on the written test, when successful and partially successful responses were combined and unsuccessful and not modified responses on the written test were combined, it was found that students produced significantly more successful and partially successful responses for +recast episodes than for –recast episodes. On the stimulated correction task, it was also found that when students did produce successful modifications, they were significantly more likely to be modified in the same way as the recast had been provided the previous day. Thus, the results from the error correction task point to at least some benefit from the recasts.

A comparison of the error correction results from the stimulated correction task and the written test showed that overall the students performed better on the written test, but that there were greater improvements for the +recast episodes than the –recast episodes, both in an increase in number of successful and partially successful modifications, and in a decrease in the number of unsuccessful modifications. It was also found that when students increased the number of successful and partially successful modifications for +recast episodes, over three quarters of these were done in the same way as the recasts had been provided. These improvements for +recast episodes reinforce that the recasts were likely of some benefit to the students in terms of their ability to modify their utterances.

For correct episodes, it was found that when students identified them as containing an error and modified them, they sometimes produced a modification that was also grammatically correct. However, there were also a number of instances in which

students changed the originally correct utterances into incorrect utterances, indicating that the students often may not be aware of when they are producing grammatically correct utterances and when they are producing grammatically incorrect utterances.

4.4 Reaction Time Results

Reaction time was measured on the error detection portion of the stimulated correction task (i.e. time between the end of the video and when the student pressed the “error” or the “no error” button). Table 4.14 displays the means (in seconds), standard deviations, minimum values, and maximum values for the reaction times in detecting errors. The results show that of the three types of episodes, students were fastest on the correct episodes (2.16076 sec.). Students were slightly faster on the +recast episodes (2.24956 sec.) than on the –recast episodes (2.28148 sec.). However, a one-way ANOVA was performed and it was found that the difference in reaction times between the three episode types was not significant [$F(2, 330) = .903, p = .406$]. As the results in table 4.15 show, there was quite a bit of variation in reaction times within each episode type: from 0.96560 to 4.65070 sec. for +recast episodes, from 1.20420 to 5.40670 sec. for –recast episodes, and from 0.91940 to 4.90460 sec. for correct episodes. The fact that the variation in reaction times within each episode type was quite large may explain why the difference in reaction times between the three types of episodes was not significant.

Table 4.14. Reaction times in error detection on the stimulated correction task.

	Mean*	SD	Minimum value	Maximum value
+recast	2.24956	0.66770	0.96560	4.65070
-recast	2.28148	0.70665	1.20420	5.40670
correct	2.16076	0.70545	0.91940	4.90460

p = *ns*

*Note. The mean reaction times are presented in seconds.

Although not significant, the reaction time results still provide some interesting information. A faster reaction time is often equated with easier/quicker accessibility within the mental stores of knowledge. Thus, the faster reaction time for the correct episodes could be because students may have automatized those forms. Slightly faster reaction times for +recast episodes than for –recast episodes may be due to the fact the forms were “highlighted” by the recasts the previous day, making them more salient and more activated in the learner’s mental stores.

In addition to examining the mean reaction time for each episode type, reaction times for +recast and –recast episodes were examined in relation to students’ modifications of their errors. These results are presented in table 4.15. The reaction times for +recast and –recast episodes were very similar when the student produced unsuccessful modification or no modification (2.24476 sec. and 2.24373 sec. respectively). As for when the students produced successful and partially successful modifications their utterance, they were fastest on +recast episodes for which they modified their utterance in the same way as the recast had been provided the previous day (2.14991 sec.). On the other hand, they were slower when they modified their utterances in a different way than the recast (2.51741 sec.). They were also slower for –recast episodes (2.40696 sec.). These results suggest that the recasts may have only led to faster reaction times for some of the episodes, namely those in which the students used the

information from the recast to modify their utterance during the error correction task. It is possible that for those utterances, the recasts were particularly salient to the learner (and thus effective), and when this was the case, the student could quickly notice the error in their speech and correct it using the information that had been provided in the recast the previous day. However, a one-way ANOVA revealed that the differences in reaction times for +recast/same as recast, +recast/different than recast, and –recast episodes were not statistically significant [$F(2, 68) = 1.141, p = .326$].

Table 4.15. Mean reaction times by nature of modification for +recast and –recast episodes.

	successful & partial		unsuccessful & not modified
	same as recast	different than recast	
+recast	2.14991	2.51741	2.24476
-recast		2.40696	2.24373
	<i>p = ns</i>		

In answer to the research question of whether there are any differences between reaction times in detecting errors that received recasts and errors that did not receive recasts, it was found that there was some evidence that recasts may have helped students respond more quickly, especially when they went on to modify their utterance using the information they received in the recast, although the difference did not reach significance.

4.5 Summary of Results

When the results of +recast and –recast episodes were compared, a benefit for recasts was demonstrated in a number of ways. On the error detection task, there was a trend for students to detect more errors on the +recast episodes than on the –recast

episodes. This trend was more prominent on the stimulated correction task and less evident, but still present, on the written test. On the error correction task, no statistically significant differences were found between +recast and –recast episodes when the analysis was carried out with each type of modification (successful, partial, unsuccessful, and not modified) as a separate category. However, trends were again present that showed that students produced more successful and partially successful modifications and fewer unsuccessful and not modified responses for +recast episodes than for –recast episodes. These trends were present on both the stimulated correction task and the written test, but were most prominent on the written test. In fact, when the categories of successful was combined with partially successful and unsuccessful was combined with not modified, the results of the written test did show significant differences between the +recast and the –recast episodes; the +recast episodes led to significantly more successful and partially successful modifications than the –recast episodes, pointing to a benefit from recasts.

An additional indication of benefit from recasts came with the examination of the same as recast and different than recast modifications on the +recast episodes. On the stimulated correction task, when students produced successful modifications, they were significantly more likely to correct their errors in the same way as the recast had. This demonstrates that the recasts had some benefit because the results showed that students used the information in the recasts to make their modifications.

In addition, support for the beneficial role of recasts came with the comparison of the stimulated correction task results with the written test results. It was found that on the written test, students were able to reduce their production of unsuccessful modifications

for +recast episodes by 12.9%, but they increased their production of unsuccessful modifications by 1.3% for –recast episodes. In addition, students were able to increase their production of successful modifications on the written test by more for the +recast episodes than for the –recast episodes. Finally, when the students did increase their production of successful and partially successful modifications for the +recast episodes on the written test, the majority of those modifications were done in the same way as the recast had been provided.

Although evidence for the benefit from recasts can be seen in the trends and significant findings of the error detection and error correction tasks, the results of the reaction times did not produce significant differences between any of the episode types. There seemed to be a trend for students to react more quickly for +recast episodes when they did go on to modify those utterances in the same way as the recast had been produced, but these differences were not significant.

In sum, there does seem to be some benefit from recasts, but the benefit was not robust enough to produce significant findings on all of the measures. Specifically, while reaction time and error detection rates did not produce significant differences between +recast and –recast episodes, error correction patterns did. Possible reasons for these findings and their implications will be discussed in the next chapter.

Chapter 5 – Discussion and Conclusions

5.1 Discussion of Results

The aim of the present study was to investigate whether incidental, extensive recasts provided by a teacher in a small-group outside a classroom were beneficial to adult ESL learners. The study was guided by the four research questions presented in chapter two. In the following section, the results are discussed with reference to these questions.

1. Are errors that receive recasts detected more often than errors that do not receive recasts:
 - (a) on the stimulated correction task?
 - (b) on the written test?

The results of the error detection task showed that students perceived a large number of the +recast and –recast episodes to contain error(s) on the stimulated correction task (76.6% and 71.2% respectively) and the written test (83.8% and 82.0% respectively). The results showed a trend for students to detect more errors in the +recast episodes than in the –recast episodes. This was true for both the stimulated correction task and the written test, but the trend was slightly larger on the stimulated correction task than on the written test. However, the difference between error detection rates on +recast and –recast episodes failed to reach statistical significance on either test.

It is encouraging that there was a trend for students to detect more errors in +recast episodes than in –recast episodes; this may point to a benefit from recasts in terms of students' ability to perceive errors in their own speech. Such a trend would be logical given that, as discussed in chapter two (section 2.3.2), it has been argued that recasts may provide negative evidence to learners (see Ellis & Sheen, 2006). If in fact the recasts provided negative evidence, we would expect students to become more aware of the fact that their utterances that received recasts were ungrammatical. Thus, the trend for students to detect more errors in their +recast episodes suggests that the students benefited from the recasts, and this benefit may be linked to the negative evidence that recasts provide.

Despite the trend for a benefit from recasts, we must also consider the fact that the difference in error detection rates between +recast and –recast episodes was not significant. There are several possible explanations for why the difference was not significant. First of all, it is possible that the recasts had no impact on the learners' perception of the presence of errors in their own speech and that the trend for students to detect more errors in +recast episodes than –recast episodes was coincidence. However, the fact that the results for all of the other research questions also point to a benefit from recasts makes this possibility less likely. Also, there seemed to be an overall bias for students to think that their speech contained errors, and this bias was found across all three episode types. This bias may have been partially responsible for the high error detection rates on the +recast and –recast episodes. These high rates, all over 70%, may have been close to ceiling, thus muting any beneficial effect from the recasts and leading to the non-significant difference between the +recast and –recast episodes.

Although previous research has examined the role of recasts, it has not examined recasts in relation to how accurate students are at detecting errors in their own speech. Nonetheless, a couple of recast studies have examined students' ability to detect errors in (written) sentences. Ellis et al.'s (2006) recast study tested students' ability to detect errors in written sentences (which were not based on the student's own speech) and found that students in the recast group were able to correctly detect 84.4% of the errors in ungrammatical sentences on the immediate post-test. This result is similar to the written test error detection results of the present study. Nabei and Swain (2002) conducted a case study on an adult ESL student and found that the student was able to correctly detect errors in 56% of the sentences that had received recasts during classroom interaction. While this result is much lower than the results of the present study, this may be because the sentences the student in Nabei and Swain was asked to judge were not all based on the students' own speech; the majority (21 of the 27) were based on the speech of other students. In addition, as Nabei and Swain examined the results of only one student, it can be argued that that particular student's results may be atypical. Hopefully in the future more studies will examine how accurate students are at detecting errors in their own speech to allow for more meaningful comparisons with the results of the present study.

2. Are errors that receive recasts corrected more often than errors that do not receive recasts:

- (a) on the stimulated correction task?
- (b) on the written test?

On the stimulated correction task, students were able to successfully and partially successfully modify more of their errors from the +recast episodes (48.2%) than from the –recast episodes (38.0%). On the written test, students were also able to successfully and partially successfully modify more of their errors from the +recast episodes (72.0%) than from the –recast episodes (54.9%). While the difference in modification patterns was a trend in the stimulated correction task, it produced a statistically significant difference on the written test.

The findings from the error modification task show that recasts benefit students' ability to correct errors in their own speech. These results are especially encouraging given that the recasts that students received in the present study were spontaneous and extensive. In other words, students demonstrated an increased ability to successfully and partially successfully modify errors after having received only one recast for each linguistic form.

Although no previous studies have compared the results of post-tests for +recast and –recast episodes, several studies have examined student performance on test questions based on +recast episodes. The stimulated correction task in the present study produced results for +recast episodes very similar to the immediate post-test results of these studies. The present study found that 48.8% of errors that received recasts were successful and partially successful modified, Loewen and Philp (2006) reported an accuracy rate of 50% on test questions based on +recast episodes, and Nassaji (2006) reported that 48.8% of errors that received recasts were either successfully or partially successfully modified. The fact that these three studies all found similar results on the immediate post-tests is encouraging for the validity of the stimulated correction scores

obtained in the present study. It also reinforces the finding in the present study that the recasts were of benefit to students in terms of their ability to modify their errors.

While the stimulated correction task produced results consistent with previous studies, the results of the written test were much higher (72.0%) than Loewen and Philp (2006) (50%) and Nassaji (2006) (48.8%). This may be due to several possibilities. First, a practice effect in the present study may have led to the high scores (this is discussed further in the limitations section below). Second, the fact that students were permitted to think about the questions for as long as they wanted on the written test (which was not the case in Loewen & Philp), likely allowed the students to make use of both implicit and explicit knowledge, thus producing higher scores. Finally, these results may be higher than those of Nassaji (2006) as the written test questions in the present study consisted of isolated sentences, perhaps causing students to examine and question each sentence more closely than they had in the paragraph format used in Nassaji (2006).

Besides Loewen and Philp (2006) and Nassaji (2006), the only other studies that have employed pre-/post-tests to examine the effectiveness of recasts have involved the use of pre-selected linguistic target forms. The post-test results of these studies have varied greatly, for example Ellis, et al. (2006) found immediate post-test results for the recast group varying from 36.1% on the imitation test to 83.9% on the grammaticality judgment test. Leeman (2003) found different results for gender agreement (57.4%) and number agreement (86.96%). Ammar and Spada (2006) found relatively high scores, with 62.2% accuracy on the written test and 74.9% on the oral test. Lyster's (2004) results were much lower, and varied from 24.6% on the text-completion test to 33.1% on the binary-choice test. The error modification results of the present study fall in the middle of

all of these results, with some studies reporting much lower scores and others reporting even higher scores. The likely reason for this great amount of variation is that each of these previous studies examined different pre-selected linguistic targets. Thus, it is difficult to draw any truly meaningful conclusions from the comparison of these results with the results of the present study, which examined extensive recasts.

The results from the error modification task clearly show a positive relationship between recasts and students' ability to correct their own errors. In addition, there is evidence that the specific content of the recasts may have assisted students in the modifications of their errors.

3. Are there any differences in students' reaction times in detecting errors for episodes that receive recasts and those that do not?

The results show that students responded slightly more quickly for +recast episodes than for –recasts, but this difference was not significant. The fact that students were able to respond more quickly indicates that students were possibly more confident (either consciously or unconsciously) in their answers for +recast episodes than for –recast episodes. While this trend did not reach significance, this was likely due at least in part to the great amount of variability in reaction times between students.

There was also a trend for students to respond more quickly in the error detection task if they later went on to provide a successful or partially successful modification that was the same as the recast they had been given during the interaction. Although students were instructed to first decide if their utterance contained any errors (and to do this as

quickly as possible) and then go on to correct the error after they had made this decision, it is possible that some students mentally completed both tasks simultaneously. Evidence for this comes from the fact that some students mumbled modifications of their utterances to themselves even before they pressed the “error” or the “no error” button. The fact that some of the students may have mentally completed the two tasks at the same time may have been one of the factors that led to the great degree of variability in reaction times between students. If some students were in fact considering the modification of the utterance during the timed error detection portion of the task, then the faster reaction times when students corrected their errors in the same way as the recast could be seen as evidence that the recasts was influencing their reaction times, assisting them in detecting and correcting the errors more quickly.

It makes sense that students would be able to detect and/or correct their errors more quickly if the correct modification (provided in the recast) was easily retrievable in their mental stores. It is possible that the recast provided the day before was stored in a location that was “easily accessible”, at least for a short period of time after it was provided (such as the one day between the interaction and the test). Since the recast was more easily accessible than other correct forms, the student was able to respond more quickly when using the modification provided in the recast as opposed to a modification of their own devise. It is also possible that when a student receives corrective feedback (such as recasts), these corrections are processed in different manner than positive evidence alone, leading to a different type of mental representation of forms that are recast (perhaps a highlighted form) than of forms that are simply modeled. As such, this

“highlighted” form may be easily accessible, allowing students to complete the error detection task/modification task more quickly.

4. Does the type of test make a difference in the results (i.e. stimulated correction vs. written test)?

There were a number of differences between the results of the stimulated correction task and the written test. On the error detection task, students detected more errors on the written test than on the stimulated correction task. This led to more accurate responses for the +recast and –recast episodes, but less accurate responses for the correct episodes. These differences point towards a bias for students to answer “error” more often on the written test than on the stimulated correction test. It is possible that the longer the students had to think about their utterances, the more they began to doubt the grammaticality of their utterances, leading to more “error” responses. This would suggest that the results of the stimulated correction task may provide a more accurate picture of students ability to detect errors in their own speech than the written test. Another possible reason for the discrepancy in results between in the stimulated correction task and the written test is that when students complete written tests, they may use more metalinguistic knowledge to make their judgments, whereas for oral tests they may rely more on instinctual feelings about the grammaticality of the utterances. The use of different types of knowledge for the different tests could lead to different results.

The error correction task also resulted in different results between the stimulated correction task and the written test. While overall, students performed more accurately on

the written test than on the stimulated correction task, the difference in scores between the two tests was not equal across the three episode types, indicating that more than a practice effect was likely involved in the differences in scores from the stimulated correction task to the written test. The results showed that students increased the accuracy of their error modification performance for the +recast episodes more than for the –recast episodes. This, coupled with the fact that the difference in error modification scores between +recast and –recast episodes was significant on the written test, but only a trend on the stimulated correction task, indicates that the written tests may have demonstrate the benefit from recasts more clearly than the stimulated correction task. This would be consistent with previous studies that have shown that more benefit for recasts (compared with no feedback/control groups) was shown on written tests than on oral tests (Ammar & Spada, 2006; Han, 2002; Lyster, 2004).

On the other hand, there was one study that found results conflicting with the present study's findings. Ellis et al. (2006) found that implicit tests (oral) showed more benefit from recasts than explicit tests (written). This difference in results may have been caused by the linguistic simplicity of the pre-selected target form (the past tense –ed) in Ellis et al. as compared with the extensive range of errors that received recasts in the present study. It is quite possible that the students in Ellis et al. already had high levels of explicit knowledge about the past tense –ed (as demonstrated by their high written pre-test average score of 78.4%). This may have meant that any effect from the recasts would have benefited their implicit knowledge levels, but not their (already high) explicit knowledge levels. In the present study, however, the students may have had lower levels

of both explicit and implicit knowledge about the forms, thus allowing a benefit from recasts to be demonstrated on both the written and the oral tests.

As suggested in the previous chapter, the greater demonstrated benefit of recasts on the written test than on the stimulated correction task in the present study may have been because the students had more time to complete the written test. While there was still some benefit from recasts demonstrated in the stimulated correction task, it is possible that the 8 seconds provided to correct the errors was not enough time for students to retrieve and implement the correct form from their memory of the previous day's interaction. However, the extra time of the written test may have allowed them enough time to retrieve the information that they learned from the recasts.

Altogether, the results of both the stimulated correction task and the written test point towards a benefit from recasts. The results also showed that this benefit is more evident in the written test. This provides evidence for the claim that recasts provided during oral interaction may also benefit students' ability to correct errors in the written mode. It also suggests that recasts, which are often considered to be an implicit form of corrective feedback (but this is in debate; see Ellis, et al., 2006 for discussion), may be able to benefit students' explicit knowledge.

This study also examined how students responded to correct episodes and how these responses compared to the +recast and –recast episodes. The data from the correct episodes showed that students perceived a considerable portion of their correct episodes to contain errors. On the stimulated correction task, students indicated that 47.7% of the correct episodes contained errors, while the rate for the written test was even higher, at 53.2%. These results indicate that students are often wrong in the assessment of the

grammaticality of their own utterances, with a tendency to think that their utterances are less grammatical than they are. Results also showed that students often modified their grammatical utterances in correct episodes in such a way that they became ungrammatical.

The results of the correct episodes may stem from a lack of confidence on the part of students about their English skills. During the interaction and the testing, several students made comments such as, “Sorry for my English,” “My English is bad,” and “I make so many mistakes.” Such comments indicated that these students felt that they had made/would make many mistakes in English. In addition, when I reviewed the written test results with the students, many seemed surprised when I indicated to them that some of the utterances they had marked as erroneous were in fact correct. Additionally, the finding that students often changed their responses for correct episodes from the stimulated correction task to the written test signals that the students may not have been confident in their responses. Thus, students’ high levels of uncertainty about their own speech likely played a role in these low accuracy scores for correct episodes.

It is also possible that the provision of recasts during the interaction may have increased students’ belief that they were frequently producing erroneous utterances. Since all of the students received a number of recasts during the interaction, they may have assumed that they were producing quite a few errors in their speech. This may be particularly true if the students from classrooms where corrective feedback is used infrequently. If this was the case, it could explain why students answered “error” so frequently on the error detection task. It would also signify that recasts may produce effects on students beyond the error targeted by the recast. Specifically, the provision of

recasts may also push students to question the grammaticality of their utterances that did not receive recasts. This possibility has yet to be examined by other researchers. In general, recast studies have only examined the effect of recasts on those individual erroneous utterances that received recasts (Loewen, 2005; Loewen & Philp, 2006; Nassaji, 2006, 2007) or on a target linguistic form (Ammar & Spada, 2006; Ellis et al., 2006; Ishida, 2004; Lyster 2004; Mackey & Philp, 1998). As far as I know, no studies have examined the effect of recasts on students' perception of their own utterances that did not receive recasts (i.e correct utterances and erroneous utterances that did not receive recasts). Ellis, et al. (2006) did examine students' ability to detect errors in both grammatical and ungrammatical sentences and found that students mistakenly thought that the grammatical sentences were erroneous 16.7% of the time, which was much lower than rates found in the present study. However, this was likely due to the fact that the students in that study were not examining their own speech, but rather sentences created by the researcher. As such, the issues of confidence in their own speaking ability and the effect of recasts on students' perception of all of their utterances would not have played a role in Ellis et al.'s findings.

A final possible cause of the high number of perceived errors in correct episodes is that the nature of the testing situation may have pushed students to frequently answer "error". As I did not indicate to students before the test began what I was researching specifically, some students may have invented for themselves a target of the testing. Some of these students may have assumed that I would be examining erroneous utterances only (after all, why would I ask students to detect errors in and correct utterances that were correct to begin with?). As such, students may have assumed that all

(or most) of the utterances were erroneous, even though I explicitly told the students that some of their utterances were correct.

5.2 Conclusions

All of the results of the present study, including those that reached significance and those that did not, point in the same direction: Recasts were of benefit to students in terms of their ability to detect and correct errors in their own speech. Specifically, when presented with errors in their own speech in video clips and in written form, students were able to detect, as well as successfully and partially successfully modify, more of the errors that had received recasts during the interaction than those that had not received recasts. In addition, there is some evidence that recasts allowed students to detect and correct their errors more quickly. Thus, overall, it can be concluded that the recasts in the present study were beneficial to students.

The positive results of this study are particularly noteworthy in two respects. First, the recasts were beneficial even though they were provided in small-group (rather than dyadic) interaction. This indicates that recasts can be beneficial even when the teacher's attention is divided between several students. Secondly, the results of the present study show that recasts can be beneficial to students even when they are provided incidentally and extensively. Thus, while previous studies have demonstrated a benefit from intensive recasts (Ammar & Spada, 2006; Doughty & Varela, 1998; Ellis et al., 2006; Leeman, 2003; Lyster, 2004; Mackey, 2006; Mackey & Philp, 1998; McDonough & Mackey, 2006; Philp, 2003, among others), the findings of the present study demonstrate that recasts do not necessarily need to be provided intensively to be effective; even a single

recast can be of benefit to students. These results confirm Loewen and Philp's (2006) findings that spontaneous and extensive recasts can benefit students. As teachers often naturally provide incidental, extensive recasts in their classes, these results should be seen as especially encouraging for both researchers and teachers.

Finally, the present study also found evidence that students often perceive their own speech to be erroneous when in fact it is grammatical. This finding is noteworthy as it raises concerns regarding student confidence levels and the accuracy of student perception of the grammaticality of their own speech. It indicates that students' perception of their own utterances (both grammatical and ungrammatical) needs to be examined further.

5.3 Pedagogical Implications

The findings of the present study have several pedagogical implications. First, the results should be reassuring to teachers who currently employ spontaneous, extensive recasts in their adult L2 classrooms. Previous studies that have examined intensive recasts may have made teachers feel that recasts could only be beneficial if they were provided intensively. The present study shows that recasts can be effective when provided in response to a wide range of linguistic errors, even if some linguistic forms receive only one recast. As such, teachers should not be discouraged from incorporating spontaneous, extensive recasts into communicative-based oral interaction with their students.

In addition, the benefit of recasts demonstrated in the present study provides motivation for the inclusion of instruction on recasts in teacher training programs.

Specifically, teachers-in-training should be made aware of what recasts are, their benefits to students, and how they can be incorporated into meaning-based student-teacher interaction in order to achieve focus-on-form goals within the classroom.

Finally, the findings from the correct episodes indicate that students often think that their correct utterances are ungrammatical. Perhaps more attention should be given by teachers to grammatical utterances; teachers should use positive-reinforcement to let students know when they are producing grammatical utterances. This could take the form of either verbal cues (“Correct”, “Well expressed”, etc.) or non-verbal cues (thumbs up, nodding, etc.).

5.4 Contributions to Research Design

The present study has incorporated a number of innovative methodological features. First, in this study each student was tested on episodes that involved errors in their speech that did receive recasts, as well as episodes that involved errors that did not receive any type of feedback (were ignored). While a within-subject design may not be necessary for studies of recasts directed towards pre-selected target forms, such methodology may be very important in the study of spontaneous, extensive recasts. Unfortunately, previous post-test studies of spontaneous, extensive recasts (Loewen, 2005; Loewen & Philp, 2006) have not compared their test scores for errors that received recasts with anything. Thus, the method in the present study, which compares student responses on errors that did receive recasts with their responses on error that did not receive recasts, makes an important methodological contribution to the study of spontaneous, extensive recasts.

A second innovation of the present study is the use of *stimulated correction* as an instrument to measure learning. Stimulated correction was designed by the researcher and, to the best of my knowledge, is the first individualized post-test to make use of video clips in the assessment of language learning. This has several advantages. First, it allows students to see and hear the exact context in which the errors were made. Second, it forces students to make their judgments quickly, as they are only permitted to watch the video clip once. This hopefully increases the chances that students are making use of their implicit knowledge to complete the task. This may be important given that it is often assumed that recasts as an implicit form of corrective feedback (Ellis et al., 2006).

A final significant contribution of the present study is that it was the first study in the area of recasts to examine student reaction time when completing post-tests. While the reaction time results of the present study did not produce significant findings, I do feel that the measurement of reaction time, which has led to significant advancements in knowledge in fields such as psycholinguistics, is very relevant to the area of corrective feedback and will lead to important advancements in knowledge in this area in the future.

5.5 Limitations and Implications for Future Research

Despite the contributions of the present study to the field of recast research in terms of research findings, pedagogical implications, and research design, there are a number of limitations to the present study. First, the number of students who participated in the present study was relatively small. With 26 participants, 23 of whom completed all tasks, it is difficult to draw firm conclusions from the findings. In addition, these small numbers may have been (at least partially) responsible for the lack of significant findings

on some of the measures despite clear trends for a benefit from recasts on all of the measures. As such, a replication of the present study with a larger sample size should be carried out. A larger sample size would hopefully lead to significant findings.

Another limitation of this study is that it was conducted in a small-group environment (as opposed to a classroom environment). While this likely reflected a much more natural situation than that of dyadic studies, we must keep in mind that the results cannot necessarily be applied to a whole-class situation. Specifically, it is likely that a student in this study received more recasts during the interaction than they would have received during a similar amount of time of whole-class activities in an actual L2 classroom. This being said, it should be kept in mind that within L2 classrooms, teachers often divide students into small groups to complete tasks, projects, etc. As such, the results of this study may be applicable to such contexts within L2 classrooms. Nonetheless, it would be worthwhile to apply the stimulated correction methodology established in this study to the study of recasts in the classroom context. Specifically, classroom interaction could be video-recorded and the videotapes used to create clips for the students to watch and complete the error detection and error correction tasks. Stimulated correction could be applied in the classroom context for both observational studies in which the teacher is not told to provide any particular type of feedback and studies in which the type and amount of feedback are manipulated by the researcher. For such studies, it should be kept in mind that high-quality sound and video are required for the play-back to students during the testing phase. Ideally, several cameras and microphones could be used to capture this data in the classroom. Alternatively, one subgroup of students from a class could be chosen for study within the classroom context.

A second limitation regarding the small-groups is the fact that the small groups examined in the present study were not all of the same size. Due to variations in student turn-out, the small groups ranged in size from three to five students. This difference could have led to slightly different effects in each of the groups, thus increasing the variation in responses, and decreasing the likelihood of significant results. However, could be argued that this difference in the size of the groups actually reflects the situation in real L2 classrooms more closely than if all of the groups had been of uniform size. When students are divided into small groups for activities in L2 classrooms, several factors (number of students in the class, student absenteeism, nature of the activity, teacher preference for group sizes, etc.) would lead to groups not always being of uniform sizes.

Another limitation of the study was the fact that all the students completed the stimulated correction task before the written test. The reason for this was that the primary goal of this study was to examine students' responses on the new methodology, stimulated correction. The written test was administered in order to have an additional measure with which to compare the results of the stimulated correction task, thus increasing validity. However, since the number of students in the present study was not large, it was decided to not further divide the students into two groups to receive the stimulated correction task and the written test in a cross-balanced manner. This means that a practice effect likely affected the results of the written test. Still, it should be remembered that the purpose of the test was not to see how accurate students were on +recast episodes, but to compare how accurate they were on +recast and –recast episodes. It was assumed that any practice effect would equally affect the +recast and –recast episodes, and not affect the comparison of these results. In addition, since the recasts

were cut out of the video clips, the possibility that students would learn from hearing the recasts a second time was avoided. In future studies, a larger sample size would allow for cross-balancing the order of testing, with half of the students receiving the stimulated correction task first and half receiving the written test first.

Finally, there remain a number of limitations relating to the applicability of these results to other situations and contexts. For example, not all types of recasts were examined in the present study. No conclusions can be drawn regarding the benefits of any types of recasts that do not fall under the definition of recasts adopted in this study. In order to examine these, the study could be replicated with other types of recasts. In addition, other types of FoF, both reactive and preemptive, could also be examined using stimulated correction. For examining preemptive FoF, the video clips could be cut immediately after the student-/teacher-initiated question and the student's task when viewing the video would be to answer the question as best they can. The undertaking of such studies could also be expanded to examine different L2s, different school environments, and students with different levels of language proficiency.

Finally, some preliminary findings of the present study should be investigated in further detail. The results of the present study raised the idea that recasts may affect students' perception of not only those errors that received recasts (+recast episodes), but also those errors that did not receive recasts (-recast episodes) and utterances that contained no errors (correct episodes). In order to examine the effect that recasts have on students' perception of -recast and correct episodes, the present study could be replicated, with a treatment group completing all the tasks in the same manner as in the present study and a control group of students completing all tasks, but not receiving any

recasts during the small-group interaction. The results of the two groups would be compared to see if there were any differences in students' perceptions of, error detection rates in, and error modification patterns in –recast and correct episodes. The results of the present study also raised the possibility that students have a bias to think their speech contains errors, but that some students were much more likely to respond “error” than others. It would be interesting to examine if there is a relationship between students' confidence in their English skills and their level of error detection in +recast, -recast, and correct episodes. A survey, asking questions that would assess student confidence in English skills, could be administered prior to the interaction in which the students receive recasts. After all of the testing, the students could be placed into groups based on their confidence levels, and then the students' error detection and correction scores could be compared to see if (a) there is any difference between the groups in terms of their scores, and (b) if there is any difference between the groups in terms of how the recasts affected their responses (comparison between +recast and –recast episodes).

The present study has answered the research questions it set out to address, but, as can be seen, it has also initiated the discussion of several new questions. Therefore, a great deal more of research is still needed in this area.

REFERENCES

- Ammar, A., & Spada, N. (2006). One size fits all? Recasts, prompts, and L2 learning. *Studies in Second Language Acquisition*, 28, 543-74.
- Braidi, S. (2002). Reexamining the role of recasts in native-speaker/non-native speaker interactions. *Language Learning*, 52, 1-42.
- Carpenter, H., Leon, K., MacGregor, D., & Mackey, A. (2006). Learners' interpretations of recasts. *Studies in Second Language Learning*, 28, 209-36.
- Crookes, G., & Gass, S. (Eds.). (1993). *Tasks and language learning*. Clevedon, U.K.: Multilingual Matters.
- Doughty, C., & Pica, T. (1986). 'Information gap tasks', An aid to second language acquisition? *TESOL Quarterly*, 20, 305-25.
- Doughty, C., & Varela, E. (1998). Communicative focus on form. In C. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 114-138). New York: Cambridge University Press.
- Doughty, C., & Williams, J. (Eds.) (1998). *Focus on form in classroom second language acquisition*. New York: Cambridge University Press.
- Ellis, R. (1994). Uptake as language awareness. *Language Awareness*, 4, 147-160.
- Ellis, R., Basturkmen, H., & Loewen, S. (2001a). Learner uptake in communicative ESL lessons. *Language Learning*, 51, 281-318.
- Ellis, R., Loewen, S., & Erlam, R. (2006). Implicit and explicit corrective feedback and the acquisition of L2 grammar. *Studies in Second Language Acquisition*, 28, 339-68.

- Ellis, R., & Sheen, Y. (2006). Reexamining the role of recasts in second language acquisition. *Studies in Second Language Acquisition*, 28, 575-600.
- Farrar, M. (1990). Discourse and the acquisition of grammatical morphemes. *Journal of Child Language*, 17, 607-624.
- Gass, S., & Varonis, E. (1994). Input, interaction, and second language production. *Studies in Second Language Acquisition*, 16, 283-302.
- Han, Z. (2002). A study of the impact of recasts on tense consistency in L2 output. *TESOL Quarterly*, 36, 543-72.
- Han, Y., & Ellis, R. (1998). Implicit knowledge, explicit knowledge and general language proficiency. *Language Teaching Research*, 2, 1-23.
- Havranek, G. (1999). The effectiveness of corrective feedback: Preliminary results of an empirical study. *Acquisition et Interaction en Langue Étrangère*, 2, 189-206.
- Ishida, M. (2004). Effects of recasts on the acquisition of the aspectual form *-te i-(ru)* by learners of Japanese as a foreign language. *Language Learning*, 54, 311-94.
- Iwashita, N. (2003). Negative feedback and positive evidence in task-based interaction: Differential effects on L2 development. *Studies in Second Language Acquisition*, 25, 1-36.
- Johnson, K. (1981). The four skills in communicative language teaching: Writing. In K. Johnson, & K. Morrow (Eds.), *Communication in the classroom*. London: Longman.
- Krashen, S. (1982). *Principles and practice in second language acquisition*. Oxford: Pergamon.

- Leeman, J. (2003). Recasts and second language development. *Studies in Second Language Acquisition, 25*, 37-63.
- Lightbown, P., & Spada, N. (2001). Review article: Recasts as feedback to language learners. *Language Learning, 51*, 719-758.
- Lochtman, K. (2002). Oral corrective feedback in the foreign language classroom: How it affects interaction in analytic foreign language teaching. *International Journal of Educational Research, 37*, 271-83.
- Loewen, S. (2005). Incidental focus on form and second language learning. *Studies in Second Language Acquisition, 27*, 361-86.
- Loewen, S., & Philp, J. (2006). Recasts in the *adult* English L2 classroom: Characteristics, explicitness, and effectiveness. *The Modern Language Journal, 90*, 536-56.
- Long, M. H. (1981). Input, interaction, and second language acquisition. In H. Winitz (Ed.), *Native language and foreign language acquisition* (pp. 259-278): Annals of the New York Academy of Sciences, Vol. 379.
- Long, M. H. (1983). Native speaker/non-native speaker conversation and the negotiation of comprehensible input. *Applied Linguistics, 4*, 126-141.
- Long, M. H. (1996). The role of the linguistic environment in second language acquisition. In W. Ritchie & T. Bhatia (Eds.), *Handbook of research on second language acquisition* (pp. 413-468). New York: Academic.
- Long, M. H. (2006). *Problems in SLA*. Mahwah, N. J.: Lawrence Erlbaum Associates.

- Lyster, R. (1998). Negotiation of form, recasts, and explicit correction in relation to error types and learner repair in immersion classrooms. *Language Learning, 48*, 183-218.
- Lyster, R. (2004). Differential effects of prompts and recasts in form-focused instruction. *Studies in Second Language Acquisition, 26*, 399-432.
- Lyster, R., & Ranta, L. (1997). Corrective feedback and learner uptake: Negotiation of form in communicative classrooms. *Studies in Second Language Acquisition, 20*, 37-66.
- McDonough, K. (2005). Identifying the impact of negative feedback and learners' responses on ESL question development. *Studies in Second Language Acquisition, 27*, 79-103.
- McDonough, K., & Mackey, A. (2006). Responses to recasts: Repetitions, primed production, and linguistic development. *Language Learning, 56*, 693-720.
- Mackey, A. (2006). Feedback, noticing and instructed second language learning. *Applied Linguistics, 27*, 405-30.
- Mackey, A., & Philp, J. (1998). Conversational interaction and second language development: Recasts, responses, and red herrings? *The Modern Language Journal, 82*, 338-56.
- Nassaji, H. (2006, July). *Explicitness of interactional feedback and its role in the learning of the targeted form in student-teacher interaction*. Paper presented at the Pacific Second Language Research Forum (PacSLRF), Brisbane, Australia.
- Nassaji, H. (2007). Elicitation and reformulation and their relationship with learner repair in dyadic interaction. *Language Learning, 57*, 511-48.

- Nassaji, H. & Hawkes, L. (2006, October). *Focus-on-form in L2 classrooms and the role of language proficiency*. Paper presented at the Second Language Research Forum (SLRF), University of Washington, Seattle, WA.
- Nicholas, H., Lightbrown, P. M., & Spada, N. (2001). Recasts as feedback to language learners. *Language Learning, 51*, 719-58.
- Panova, I., & Lyster, R. (2002). Patterns of corrective feedback and uptake in an adult ESL classroom. *TESOL Quarterly, 36*, 573-95.
- 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.
- Sabbagh, S. (1998). *Learners’ reactions to feedback in an adult ESL classroom*. Unpublished master’s thesis, Pepperdine University, Malibu, C.A., USA.
- Saxton, M. (1997). The contrast theory of negative input. *Journal of Child Language, 24*, 139-161.
- Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics, 11*, 129-158.
- Schmidt, R. (Ed.), (1995). *Attention and awareness in foreign language learning*. Honolulu, Hawaii: University of Hawaii Press.
- Sheen, Y. (2006). Exploring the relationship between characteristics of recasts and learner uptake. *Language Teaching Research, 10*, 361-92.
- 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.

Williams, J. (2001). The effectiveness of spontaneous attention to form. *System*, 29, 325-240.

APPENDIX A

Transcription Conventions

T	teacher
S	student
+	pause
-	false start
<< >>	extra-lingual information (such as laughing)
♪	auditory cue
<u>Underlined</u>	erroneous student utterance
Bold	recast
<i>Italics</i>	correct student utterance
<i>Bold-italics</i>	student modification

APPENDIX B

Background Questionnaire

1. Name: _____

2. Age: _____

3. Nationality: _____

4. Native language: _____

5. What other languages can you speak and how well can you speak them?

		Very well				Not very well
i.	_____	1	2	3	4	5
ii.	_____	1	2	3	4	5
iii.	_____	1	2	3	4	5

6. How many months have you lived in Canada? _____

7. Do you live with a host family? ____ yes ____ no

8. Have you ever lived in another English speaking country? ____ yes ____ no

If yes, which country? _____

How long did you live there? _____

9. How long did you study English before coming to Canada?

_____ years and _____ months.

10. Why are you studying English?

_____.

APPENDIX C

Sample CVs

Lisa Harris

1555 Chinook Street
Victoria, BC V8G 2P9
(250) 123-4567
lharris@yahoo.ca

Profile

- ★ Three years teaching English to children in Japan
- ★ Native speaker of English
- ★ Can speak Spanish fluently
- ★ Can speak a little Japanese
- ★ Likes drawing, painting and doing art

Work Experience

- ◆ **Teaching ESL to children in Japan (2002-2005)**
Private English school
Taught three classes of children aged 4-7
- ◆ **Research Assistant (part-time) (2001-2002).**
English department, University of Victoria
- ◆ **Waitress (part-time) (2000-2001).**
Boston Pizza, Victoria, BC
- ◆ **Salesperson (part-time) (1999-2000).**
Zellers, Victoria, BC

Education

- ◆ **B.A. in English from University of Victoria (2002)**
Awards: Entrance scholarship
Courses: English literature, linguistics, Spanish, psychology
- ◆ **Teaching English as a Second Language (TESL) certificate (January-March, 2001)**
- ◆ **High School Diploma from Victoria High School (1998)**

APPENDIX D

Sample written test

Student code: _____

Date: _____

Read these sentences taken from your speech during the discussion. First decide if there is any error(s) in the sentence by putting a \checkmark beside Error or No error. Then, if there is an error, correct it by writing down the correct way to say it.

Examples:

a. Where he studied? \checkmark Error ___ No error

Where did he study?

b. Maybe Lisa is the best choice. ___ Error \checkmark No error

1. If you apply a job in a company you write a letter ___ Error ___ No error

2. CV is international name. ___ Error ___ No error

3. He or she must know everything for their language. ___ Error ___ No error

4. Everybody knows general rules. ___ Error ___ No error

5. She's from Russian. ___ Error ___ No error

6. This topic is important for her writing. ___ Error ___ No error

7. He PhD in Linguistic from UBC. ___ Error ___ No error

8. After that he researched assistant in the linguistics department in Laval University.

Error No error

9. He researched assistant in the linguistic department University of Victoria in 2003
between 2006.

Error No error

10. He has one year English teaching experience.

Error No error

11. We can't say anything - he or she must be sociable.

Error No error

12. He native speaker of French.

Error No error

13. He worked research assistant from university.

Error No error

14. He very interested in ESL research.

Error No error

15. If you can research something linguistic it's an advantage for you.

Error No error

16. They explain that differences the two languages.

Error No error

APPENDIX E

Instructions for the stimulated correction task

In a few minutes, you will watch some short sections of video from yesterday's discussion. You will watch them on this computer screen and you will listen using these headphones. Before we watch the sections of video, I will explain what I want you to do. After I explain, I will show you an example using a video of myself.

When we start, the first thing you will see is this: "The video will begin in 3 seconds"

<<computer screen shows "The video will begin in 3 seconds">>

After three seconds, the video will start to play. When the video is playing, I want you to listen carefully to your own speech. You may hear myself or some other students speaking as well, but listen especially carefully to your own speech. During the video, you will hear a sound like this <<play beep>>. I want you to listen very carefully to your own speech after the beep. I want you to decide if you think your speech after the beep had any errors or mistakes. Do not worry about pronunciation errors. Only think about other types of errors. Let's watch an example video of myself. After you watch the video, you will see "Press error or no error" on the screen. If you think your speech did have an error, you press this button <<pointing to "Error" button>>. If you think your speech was correct, did not contain any errors, you press this button <<pointing to "No error" button>>. I want you to listen, make a decision, and press the button as quickly as you can. If you are not sure if your speech had any errors, you can guess.

If you press the "No error" button, the computer will show "The video will begin in 3 seconds" and you will wait for the next video to play. If you press the "Error" button, the computer will show "Now correct it" <<press the "error" button to display the "Now correct it screen">>. I want you to correct the mistake in your speech by speaking into this microphone. Tell me the correct way to say it by speaking the whole sentence. You will have 8 seconds to do this before the next video begins. Again, if you are not sure, you can guess the correct way to say it.

Do you have any questions before I show you an example? <<answer any questions>>

Now I will show you some example sections of video of myself. Listen to my speech and watch what I do. <<run example clips>>

Do you understand what to do? Do you have any questions? Now you will do some practice ones. Just try your best and don't worry because these are only practice. Remember, listen to your speech after the beep and decide as quickly as you can if there are any errors. Do not worry about pronunciation errors. <<student completes practice clips>>

Was that okay? Do you have any questions before we begin the real ones? <<answer any questions>> Now remember, try to make your decision and press the button as quickly as you can.