

Interactional Feedback and ESL Question Development

by

Abbey Bell
B.A., University of Victoria, 2005

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

MASTER OF ARTS

in the Department of Linguistics

© Abbey Bell, 2008
University of Victoria

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

SUPERVISORY COMMITTEE

Interactional Feedback and ESL Question Development

By

Abbey Bell
B.A., University of Victoria, 2005

Dr. Hossein Nassaji (Department of Linguistics)

Supervisor

Dr. Hua Lin (Department of Linguistics)

Departmental Member

ABSTRACT

Supervisory Committee

Dr. Hossein Nassaji, Supervisor
(Department of Linguistics)

Dr. Hua Lin, Departmental Member
(Department of Linguistics)

Interactional feedback has received a lot of attention recently in the field of second language acquisition (SLA). In particular, the literature on the effects of recasts (correct reformulation of a learner's utterance) has produced conflicting results. The object of the present study is to contribute to the recent investigation of interactional feedback on L2 acquisition by examining the effects of recast, elicitation, and recast plus elicitation on the development of question formation by ESL learners in a typical classroom environment. The study followed a pre-test/post-test design and was carried out over 7 weeks. Results provide some support for short-term effects of recasts, long-term effects of elicitations and delayed effects for recasts in combination with elicitation. These results suggest that recasts and elicitation may both be effective forms of feedback in different ways. As well, the results also imply that recasts may be most effective when their saliency is enhanced in some manner.

TABLE OF CONTENTS

SUPERVISORY COMMITTEE.....	ii
ABSTRACT.....	iii
TABLE OF CONTENTS.....	iv
LIST OF TABLES.....	vii
LIST OF FIGURES.....	viii
ACKNOWLEDGEMENTS.....	ix
Chapter One INTRODUCTION.....	1
1.1 RESEARCH STATEMENT.....	1
1.2 BACKGROUND	2
1.3 RESEARCH QUESTIONS	3
1.4 RESEARCH HYPOTHESES	4
1.5 THESIS OUTLINE	5
Chapter Two LITERATURE REVIEW.....	7
2.1 COMPREHENSIBLE INPUT AND THE INTERACTION HYPOTHESIS.....	8
2.2 FOCUS ON FORM.....	9
2.3 INTERACTIONAL FEEDBACK.....	12
2.3.1 <i>Recasts And Elicitations</i>	13
2.3.2 <i>Mixed Results For Feedback Effects</i>	17
2.3.3 <i>The Challenge of Investigating Feedback</i>	24
2.3.4 <i>Research Needed</i>	32
Chapter Three METHODOLOGY	37
3.1 PARTICIPANTS.....	37
3.2 RESEARCH DESIGN.....	41
3.2.1 <i>Data Collection Procedures</i>	45
3.2.2 <i>Pre-Test</i>	45

3.2.3	<i>Treatment</i>	46
3.2.4	<i>Post-Tests</i>	48
3.3	CODING.....	49
3.3.1	<i>Developmental Stages</i>	50
3.3.2	<i>Improvement</i>	51
3.3.3	<i>Grammaticality</i>	52
3.3.4	<i>Reliability</i>	53
3.4	STATISTICAL ANALYSES	54
Chapter Four ANALYSES AND RESULTS.....		55
4.1	TREATMENT AND CONTROL GROUPS.....	55
4.1.1	<i>Overall Stages By Control And Treatment Groups</i>	56
4.1.2	<i>Improvement By Control And Treatment Groups</i>	57
4.1.3	<i>Grammaticality Of Control And Treatment Groups</i>	59
4.2	RECAST, ELICITATION AND RECAST + PROMPT	61
4.2.1	<i>Overall Stages Of Feedback Groups</i>	62
4.2.1.1	Overall Stages on Pre-test and Post-test 1	62
4.2.1.2	Overall Stages on Post-test 2	68
4.2.2	<i>Improvement Of Treatment Groups</i>	71
4.2.2.1	Improvement of Treatment Groups on Post-test 1	71
4.2.2.2	Improvement of Treatment Groups on Post-test 2.....	73
4.2.3	<i>Grammaticality Of Treatment Groups</i>	77
4.2.3.1	Grammaticality of Treatment Groups on Pre-test and Post-test 1	77
4.2.3.2	Grammaticality of Treatment Groups on Post-test 2	81
4.3	SUMMARY OF RESULTS	83
Chapter Five DISCUSSION		85
5.1	FEEDBACK EFFECTS	86
5.2	RECASTS AND ELICITATIONS	88
5.3	RECASTS, ELICITATIONS, OR RECAST + PROMPT	91

5.4	LONG-TERM FEEDBACK EFFECTS	95
Chapter Six CONCLUSION.....		100
6.1	SUMMARY OF RESEARCH.....	100
6.2	LIMITATIONS.....	102
6.3	FUTURE RESEARCH.....	103
6.4	CONTRIBUTIONS.....	105
REFERENCES.....		107
Appendix A SLEP TEST: SAMPLE QUESTIONS		112
Appendix B SAMPLE CROSSWORD		113
Appendix C SAMPLE STORY COMPLETION TASK.....		114
Appendix D BACKGROUND INFORMATION.....		115

LIST OF TABLES

TABLE 3-1 Participant Summary.....	41
TABLE 3-1 Research Schedule and Design.....	44
TABLE 3-3 Developmental Stages and Question Structures.....	50
TABLE 3-4 Sample of Coding: Grammaticality and Stage Level.....	52
TABLE 4-1 Overall Stages of Control and Treatment Groups.....	56
TABLE 4-2 Improvement of Control and Treatment on Post-test 1.....	59
TABLE 4-3 Grammaticality of Control and Treatment Groups.....	60
TABLE 4-4 Overall Stages of All Groups.....	63
TABLE 4-5 Number and Mean of Overall Stages on the Pre-test.....	64
TABLE 4-6 Mean Difference Between Groups on the Pre-test.....	66
TABLE 4-7 Mean Difference Between Feedback Groups on Post-test 1.....	68
TABLE 4-8 Overall Stages on Post-test 2.....	69
TABLE 4-9 Number and Mean of Overall Stages on Post-test 2.....	70
TABLE 4-10 Improvement of All Groups on Post-test 1.....	73
TABLE 4-11 Improvement on Post-test 2.....	75
TABLE 4-12 Grammaticality of All Groups on the Pre-test and Post-test 1.....	78
TABLE 4-13 Mean % of Grammatical Questions.....	79
TABLE 4-14 Mean Difference in Grammaticality on the Pre-test.....	80
TABLE 4-15 Mean Difference in Grammaticality on Post-test 1.....	81
TABLE 4-16 Grammaticality on Post-test 2.....	82

LIST OF FIGURES

FIGURE 1 Overall Improvement of Control and Treatment on Post-test 1.....	58
FIGURE 2 Overall Improvement of All Groups on Post-test 1.....	72
FIGURE 3 Overall Improvement on Post-test 2.....	74
FIGURE 4 Immediate and Delayed Improvement.....	76

ACKNOWLEDGEMENTS

This thesis would never have been completed if not for the help, encouragement, and guidance of many people to whom I would like to extend my deepest gratitude. I would first like to thank my supervisor, Dr. Hossein Nassaji, for believing I could do this project and for his patience as I worked through myriad frustrations and statistical turmoil. I am honoured to have had his expertise and guidance to support me through every step of the research process. In addition, I would like to thank Dr. Nassaji for contributing some financial support for my thesis and for employing me as a research assistant throughout my studies. I would also like to thank my supervisory committee member, Dr. Hua Lin, for her time and support. Thank you also to Dr. Ulf Schuetze for agreeing to be my external examiner.

A huge and grateful thank you to Carolyn Pytlyk for passing on her knowledge of everything statistical and thesis related. Another huge thank you goes out to Laura Hawkes for taking the time to slog through and code over a hundred questions. I am so fortunate to have such wonderfully supportive friends and colleagues.

I would also like to acknowledge and thank the students who participated in this research, without whom there could not have been a project. A special thanks also to the teachers who volunteered so much of their time and insight to my study.

Finally, I would like to express my warmest and most sincere gratitude to my family and friends who were and continue to be my personal cheerleading squad. Jeff, I could not have gone through all those mental brick walls without your unwavering love and support, thank you for your patience and understanding. Thank you Mom and Dad for your love and sympathetic ears, and for always believing I could do it.

Chapter One

INTRODUCTION

Interactional feedback as a means of helping second language (L2) learners to focus on form has received increased attention recently in the field of second language acquisition (SLA). In particular, recasts (correct reformulations of a learner's utterance) have been the focus of much debate (e.g., Braid, 2002; Han, 2002; Lyster & Ranta, 1997; Mackey & Oliver, 2002; Mackey & Philp, 1998; Nassaji, 2007a; Nassaji, in press). The literature investigating this area of second language pedagogy has typically described two methods of research: data are either empirically collected from teacher-student dyads where interaction is artificially altered and feedback is provided for experimental reasons, or data is collected qualitatively through classroom observation. While both methods have contributed to researchers' understanding of the relationship between feedback and second language acquisition, the former is not representative of classroom dialogue and the latter leaves too many variables open for appropriate generalizations to be made. In order to address these issues, this thesis describes a longitudinal research study that was designed to represent typical classroom activity while still controlling the type of feedback provided within each class.

1.1 Research Statement

This study examines the effects of different types of interactional feedback, specifically recast, elicitation, and recast + prompt, on the development of question formation by ESL learners in a previously established classroom environment. Thus, this

study aims to contribute to the recent investigation of different facets of interaction on L2 acquisition by isolating specific types of feedback provided within typical ESL classes.

1.2 Background

Research has shown that focusing on form may be necessary for L2 learning; interactional feedback plays an important role in helping students to notice forms when negotiating for meaning. From an interactionist perspective, conversational interaction allows for the negotiation of meaning and modifications to the interactional structure which serve to increase the comprehensibility of the input (Long & Robinson, 1998). There have been many studies that have investigated and found support for the role of negotiation and interaction in L2 learning (e.g., Lyster, 2004; Lyster & Ranta, 1997; Mackey, 1999; Mackey, Oliver, & Leeman, 2003; Mackey & Philp, 1998; Muranoi, 2000). However, there have been fewer studies comparing the effects of various *types* of interactional feedback provided within negotiations for meaning. For example, in a study investigating if and how interaction is beneficial for adult language learners' question development, Mackey (1999) only identifies one type of feedback (recasts); such studies (see also Mackey & Oliver 2002; Mackey & Philp, 1998) do not investigate which aspects of negotiation or which types of feedback are responsible for the positive effects observed.

In contrast, Lyster and Ranta (1997) identify six different types of feedback used by four French Immersion teachers in their study of corrective feedback. These feedback types were identified and categorised *after* recordings were made of regular immersion classes; feedback types were not provided in isolation so as to control for feedback

effects. However, the majority of experimental research has been conducted by and large outside the language classroom (e.g., Braidi, 2002; Mackey, Oliver, & Leeman, 2003); as such, the effects of different feedback types provided within negotiation in typical L2 classrooms remain open to investigation.

This study focuses on three types of interactional feedback: recasts, elicitations, and recasts in combination with elicitation, hereafter referred to as recast + prompt (as used by Nassaji, 2007a). Long and Robinson's (1998) operationalisation (that is, their working definition of recasts employed in their research) of recasts will be used for the purposes of the present study; as such, recasts are operationalised as corrective reformulations of a learner's utterances that do not alter the learner's intended meaning (Long & Robinson, 1998). This definition of recasts was chosen for its clarity, which renders it easily identifiable, and also for its frequent occurrence in the literature on SLA (see page 43 for an example of a recast). Elicitations (see page 42 for an example) are operationalised as utterances which indicate a need for modification without necessarily indicating that there is an error or where that error might be (although in some cases elicitation may locate the error). Recasts + prompts (see page 43 for an example) are operationalised as recasts provided immediately before or after an elicitation that (ideally) prompts the student to modify their utterance. For the purposes of this research, the location of the recast and prompt in the utterance is not examined; thus, recast + prompt refers to any feedback in which a recast and a prompt are provided in immediate succession, no matter whether the recast or the prompt occurs first. Question formation was chosen as a means of measuring the effects of these feedback types due to the developmental stages identified by Pienemann, Johnston, and Brindley (1988) through

which L2 learners of English progress. Specifically, developmental stages of question development have been used by a number of previous researchers (e.g., Mackey, 1999; Mackey & Oliver 2002; Spada & Lightbown 1993) so it will be possible to accurately compare the results of the present study to previous findings. In addition, these developmental stages are relatively easy to identify as each stage is comprised of certain components and structures, with the more complex structures occurring in the higher stages. These developmental stages make it possible to measure development in two ways: first, I examine the stage levels of questions produced by learners both prior to and after receiving feedback and second, I compare the grammatical accuracy of questions produced within a given stage.

1.3 Research Questions

The present quasi-experimental study is motivated by the challenge to control for feedback effects without affecting the spontaneity and integrity of the ESL classroom. As such, this thesis investigates the following four research questions:

1. Do learners who receive interactional feedback in the form of recast, elicitation or recast + prompt progress to higher stages of question development than learners who do not receive interactional feedback?
2. Do learners who receive recasts progress to a higher level of question development than learners who receive elicitation?
3. Is it more effective to provide recast + prompt than to provide only one of these types of feedback (only recasts or only prompts)?
4. Are positive effects of either form of interactive feedback maintained over time?

1.4 Research Hypotheses

The following hypotheses correspond to the research questions presented above.

1. Learners who receive interactional feedback (in the form of recast, elicitation or recast + prompt) will produce questions at a higher stage than learners who do not receive any feedback. This hypothesis is based on the consistent support for interaction found in the recent literature on feedback (see the literature in Chapter Two and the brief background information provided in Section 1.1 of this chapter).
2. Studies have shown that recasts typically occur quite frequently in the L2 classroom, yet the results of their effectiveness in L2 acquisition have not been conclusive. While some studies have shown recasts to be effective (e.g., Leeman, 2003 and Mackey & Philp, 1998), others have provided evidence that recasts are not as effective as elicitations and other types of feedback (Lyster & Ranta, 1997). The mixed results discovered by previous researchers make it difficult to predict which type of feedback will be more effective; thus, the second research hypothesis predicts that learners who receive recasts will not progress to a higher level of question development than learners who receive elicitation.
3. Learners who receive recast + prompt will produce questions at a higher stage than learners who receive only recasts or elicitations. It seems logical that the combination of two types of feedback into a single feedback move will increase

the likelihood that a learner will notice the form in question. In other words, the prompt may enhance the recast in some way to make it more salient.

4. Any positive effects of either form of interactive feedback will be maintained over time. Both sustained and delayed effects of feedback on question formation have been found in previous research (Mackey, 1999; Mackey & Oliver, 2002; Mackey & Philp, 1998).

1.5 Thesis Outline

As described above, the purpose of this study is to investigate feedback effects in a typical ESL classroom environment. This thesis is organised into six chapters. Chapter One provides a brief introduction to the subject of interactional feedback and presents the research statement and the questions that motivate this research. Chapter Two reviews the recent literature in the field of SLA with a focus on interactional feedback as a method of drawing learners' attention to form during meaningful interaction. A detailed description of the methodology and design of the study is provided in Chapter Three, and the analyses and results carried out in this research are presented in Chapter Four. An in-depth discussion of the results is provided in Chapter Five; this chapter also discusses how the results fit in (or disagree) with previous research. The thesis concludes by summarising the key points of the research before acknowledging the limitations of the study, suggesting avenues for future research, and outlining its possible contributions to the field of SLA in Chapter Six.

Chapter Two

LITERATURE REVIEW

This thesis aims to test the effects of three types of feedback in a quasi-experimental study set in a typical ESL classroom. Previous research on the topic of corrective feedback covers a range of feedback types, methods of data collection and observation and measures of L2 acquisition. Results of these studies are mixed and sometimes contradictory when it comes to identifying one feedback type as more effective than another. This chapter reviews the pertinent research that has investigated the role of interaction and feedback in the L2 classroom.

Each section provides a review of a different area: Section 2.1 briefly introduces the theories behind the concept of interaction as an ideal sight for language acquisition. Section 2.2 describes the research behind a “focus on form” approach to L2 classroom methodology. Section 2.3 discusses the use of interactional feedback as a way to increase learners’ awareness of certain forms. Different types of interactional feedback are examined in detail, specifically recasts and elicitations. The challenges involved in measuring the effects of different types of feedback in the L2 classroom are also examined. Section 2.3 concludes with an investigation of the noticeable dichotomy between experimental and observational research in the field and discusses the need for hybrid studies that allow researchers to control for feedback effects without interrupting the regular flow of the class.

2.1 Comprehensible Input and the Interaction Hypothesis

The term *input* comes from information processing studies¹ and in the field of SLA refers to the target language available to learners (Spada & Lightbown, 1999). In the communicative L2 classroom, where attention to meaning is the dominant focus, input is often provided through conversational interactions between the teacher and the learners, as well as between learners. Although it is widely accepted that input is crucial to L2 learning, it remains open to investigation why only some of the input is learned, and which factors determine the linguistic features or forms to be acquired. Krashen's *Comprehensible Input Hypothesis*² suggests that in order for language acquisition to occur, input must be comprehensible to the learner in order for it to be processed or learned. According to Krashen, the input available to learners must be just above their existing level of proficiency so that it provides new L2 targets without overwhelming the learners' developing interlanguage. As noted by VanPatten (1996), however, the presence of comprehensible input does not entail that all such input will be processed by the learner. Thus, VanPatten makes an important distinction between input and *intake* (VanPatten, 1996), which refers only to the linguistic data in the input that a learner actually processes.

Long's *Interaction Hypothesis* (as cited in Long & Robinson, 1998) builds on Krashen's work with the suggestion that conversational interaction is an important source of comprehensible input; such interaction between learners (and especially between learners and more proficient speakers) is a crucial site for language development. In other words, conversational interaction may help comprehensible input to become intake.

¹ Further details can be found in Sharwood Smith (1993).

² A more detailed description can be found in Brown (2000).

An important benefit of conversational interaction is the negotiations for meaning that result in “modifications to the interactional structure of conversation” (Long & Robinson, 1998). These modifications make input more comprehensible as interlocutors work through comprehension checks, clarification requests, interruptions, and confirmation checks in order to understand and to be understood. Long is not alone in his belief in the role of interaction; Swain (2000) advocates the importance of negotiation work in her *Output Hypothesis*. Comprehensible output produced within negotiations for meaning draws learners’ attention to differences between input and output, or “holes” in their interlanguage. Other studies that support the role of negotiation and interaction in L2 learning include Lyster and Ranta (1997), Mackey (1999), and Mackey, Oliver, and Leeman (2003), among others. These studies are discussed in more detail in the following two sections which explore the growing body of research investigating *focus on form* (Section 2.2) and the role of interactional feedback as a means of drawing learners’ attention to form during conversational interaction (Section 2.3).

2.2 Focus on Form

The concept of focus on form has received a great deal of attention in the field of SLA (Doughty & Varela, 1998; Ellis, 2001; Ellis, Basturkmen, & Loewen, 2001; Nassaji, 2007a). Instructional focus on form (FonF) was generated from the failure of learners to achieve high levels of fluency in communicative classrooms (Swain, 2000). There have been other conceptualisations of focus on form. For example, Lyster and Ranta (1997) claim that “negotiation of form” occurs when a teacher initiates a correction by indicating the presence of an error in a learner’s utterance, thereby providing the learner with an

opportunity to correct the error. In their definition, negotiation of form is only considered to have taken place if the learner corrects or attempts to correct the error in the immediate turn following the teacher's initiation. For the purposes of this thesis, however, FonF will be used to refer to form-focused instruction that involves planned or incidental instructional activities that are aimed at drawing learners' attention to linguistic form (Ellis, 2001). The goal of this section is to review the literature on FonF.

In his introduction to form-focused instruction research, Ellis (2001) notes that FonF originated in response to learners' failed attempts at acquiring certain grammatical forms even after years of exposure to these target structures. While communicative language teaching focuses almost exclusively on meaning, FonF includes implicit techniques aimed to “*add* attention to form to a primarily communicative task rather than to *depart* from an already communicative goal in order to discuss a linguistic feature (Doughty & Varela, 1998, p. 114).” However, it is not entirely clear which forms may be the most conducive to FonF. In a study aimed toward identifying which forms learners attended to during FonF, Poole (2005) found that the majority of the forms were lexical in nature (as opposed to grammatical, for example). Poole's study, which included 19 international students participating in an advanced ESL writing class in the United States, provided evidence that focus on form instruction may be most beneficial for vocabulary learning, but that students were unable or unwilling to focus on grammar. On the other hand, the learners in this study were highly advanced and likely chose to focus on lexical forms as a means of improving their writing. Indeed, a growing body of research has found beneficial effects of FonF on grammatical structures such as question forms, for example (Mackey, 1999; Spada & Lightbown, 1999).

Teachers can make use of either planned or incidental FonF (Ellis, 2001). While planned FonF involves tasks designed to elicit a specific structure from learners, incidental FonF involves attention to form that occurs naturally during the course of a given task (Ellis, 2001). In either case, the tasks are communicatively oriented so that attention to form arises during meaning-focused interaction (Ellis, Basturkmen, & Loewen, 2002). During either form of FonF instruction, the teacher must decide how to address or respond to problems that occur in the course of interaction.

Many studies in recent literature have focused on planned FonF: these studies include the quasi-experimental studies of Mackey and Philp (1998), Spada and Lightbown (1999), Leeman (2003), and McDonough and Mackey (2006), to name a few. Planned FonF is especially conducive to such experimental designs because it is possible to test learners' knowledge of the structure in question before any experimental treatment is provided. In addition, a researcher can structure experimental conditions in planned FonF that would not be possible in incidental FonF because of its unplanned nature (Ellis, Basturkmen, & Loewen, 2001).

A learner's attention can be drawn to problematic forms through reactive or preemptive FonF (Long & Robinson, 1998). Reactive FonF occurs when a learner (or learners) produces an erroneous utterance that is addressed by the teacher or another learner. Preemptive FonF involves the teacher or a learner drawing attention to a certain form without any production problems having occurred (Ellis, 2001). While both reactive and preemptive FonF episodes have been studied in the literature, one benefit of reactive focus on form is that it provides learners with feedback at the moment they have

something to say. This may heighten the learners' awareness of the form they misused by drawing their attention to their own language *at the time they use it*³.

Reactive FonF is optimal for investigating feedback effects for two main reasons: (1) the episode occurs after the learner has made an error, so it is clear that the structure in question is indeed problematic for the learner, and (2) the provision of implicit feedback in reaction to a communication or grammatical problem can easily be carried out as a conversational turn. With preemptive FonF, on the other hand, it is impossible to tell whether the learner would have found the structure problematic without the preemptive FonF. While Ellis, Basturkmen, and Loewen (2001) found that preemptive episodes occurred as frequently as reactive ones, they also found that they tended to involve *explicit* feedback. For these reasons, only *reactive* FonF episodes will be investigated in this thesis which focuses on implicit types of interactional feedback. The following section provides a detailed discussion on the role of interactional feedback in SLA.

2.3 Interactional Feedback

One method of implementing FonF without diverting learners' attention away from the communicative content of the classroom is to provide implicit corrective feedback during conversational interaction, or interactional feedback. Interactional feedback occurs naturally during conversation when learners and native speakers (or other learners) negotiate for meaning. Negotiation elicits feedback that may increase learners' awareness of some linguistic forms (Long & Robinson, 1998). Interactional feedback can provide the learner with positive or negative evidence; positive evidence

³ See also Nassaji, 2007b for a discussion of reactive focus on form on written errors.

provides the learner with correct information about the original utterance, while negative feedback provides evidence to the learner that something in the utterance is ungrammatical or unacceptable (Leeman, 2003). In either case, the feedback is provided *implicitly* as a conversational turn and is arguably less intrusive to the communicative goal than an *explicit* response to an error. The provision of implicit negative feedback within conversational interaction (hereafter referred to as interactional feedback) may draw learners' attention to form by making a given form more salient during a meaning-based or communicative task (Long & Robinson, 1998).

This section explores two main issues involved with measuring the effects of implicit interactional feedback. First, Section 2.3.1 describes different types of interactional feedback, namely recasts and elicitations. Section 2.3.2 explores some of the reasons behind the mixed results observed in the literature on feedback effects. Section 2.3.3 reviews some important studies that have investigated the effectiveness of each type of feedback and discusses the major difficulties involved in measuring feedback effects. Finally, Section 2.3.4 reviews the experimental and observational studies that have been done on interactional feedback and illustrates the need for more quasi-experimental studies to be carried out in the L2 classroom for conclusive results on feedback effects.

2.3.1 Recasts and Elicitations

One form of negative feedback is recasts, operationally defined by Long and Robinson (1998) as “corrective reformulations of a child’s or adult learner’s (L1 or L2) utterances that preserve the learner’s intended meaning (p. 23).” Such feedback is said to

increase the likelihood that learners will focus on the ungrammatical form in question in a way that comprehensible input alone cannot achieve. Research has shown that recasts are by far the most frequent type of interactional feedback that learners receive across a variety of language settings (e.g., Loewen & Philp, 2006; Lyster & Ranta, 1997; Panova & Lyster, 2002; Sheen, 2004). However, the focus of many recent studies has been whether this tendency to recast learner errors is also the most effective way of drawing learners' attention to form. It may be that other types of interactional feedback are more likely to be perceived as feedback on a problematic form rather than a conversational turn.

A second type of interactional feedback is elicitation, a specific type of prompt where the teacher pushes the learner to reformulate their own erroneous utterance without stating that an error has occurred or suggesting where the error may be (Lyster, 1998). The teacher's response could be interpreted as either an indication that the utterance was incorrect or as a conversational turn, such as a clarification request. For example, a teacher's response to an ill-formed utterance with a phrase such as "Can you repeat that?" might push the student to reformulate the original utterance. On the other hand, such a phrase might simply elicit a repetition of the original utterance; in this respect, prompts and recasts are similar in that they can both function as conversational turns in addition to their pedagogical functions as implicit negative feedback. Prompts also occur quite frequently in the L2 classroom, coming second only to recasts (Panova & Lyster, 2002).

Research has found some support for the use of recasts (e.g., Han, 2002; Hawkes, 2007; Nassaji, 2007c; Ohta, 2000). For example, in a small-scale study of eight upper-intermediate adult female ESL learners, Han (2002) used recasts as the only pedagogical

tool to address learners' inconsistent use of verb tense. Han's study was concerned with (a) whether learners who received recasts on their L2 output would exhibit greater tense consistency than learners who had not, and (b) whether learners who received recasts would show a higher awareness of tense consistency. All participants were randomly assigned to either a recast group or a non-recast group and were found to produce the same degree of tense consistency on a pre-test that was administered before eight instructional sessions conducted separately for each group. The results revealed a pattern of overall growth in past tense use and a decrease in mixed tenses by the recast group. These outcomes imply that recasts had a positive effect on control over tense consistency and may therefore be an effective method of providing feedback.

Conversely, other researchers have suggested that recasts are not as effective as other forms of feedback (Lyster, 2004; Lyster & Ranta, 1997). For example, Lyster and Ranta (1997) found that elicitations, metalinguistic feedback⁴, clarification requests, and repetitions lead to student-generated repairs of erroneous utterances more successfully than recasts. In contrast, another study by Lyster (2002) suggests that prompts and recasts complement each other as they have different purposes for different learners. Lyster argues that corrective feedback plays an important role in form-focused negotiation – which is argued to be “less likely to create pragmatic ambivalence than recasts embedded in meaning-focused negotiation (Lyster, 2002, p. 245).” However, Lyster does not investigate whether recasts, too, are less ambivalent in form-focused rather than meaning-focused negotiations. Indeed, in his study concerning corrective feedback and their relation to error types and learner repair in immersion classrooms,

⁴ Metalinguistic feedback is defined by Lyster and Ranta (1997) as “comments, information, or questions related to the well-formedness of the student's utterance, without explicitly providing the correct form.”

Lyster (1998a) categorised feedback types such as clarification requests and repetition of errors as negotiation of form while recasts and explicit error corrections were coded separately. In this study, Lyster suggests that negotiations of form were more effective than recasts with lexical and grammatical errors, but that recasts often led to repair of phonological errors. This finding suggests that recasts and prompts may both be effective, albeit in different ways.

While Lyster (2002) argues that recasts occur in meaning-focused rather than form-focused negotiations, it seems likely that recasts may also draw learners' attention to form in other contexts. An example of such a context might involve a task that highlights a specific form; in such tasks, teachers have an ideal opportunity to recast students intensively during classroom interaction and thereby increase the likelihood that the recast will be noticed as a focus on form, rather than meaning. While *extensive* recasts occur incidentally as problems arise, *intensive* recasts are provided repeatedly and consistently for a single linguistic item (Ellis, 2001). It may be that recasts are most effective when focused in this manner and perhaps are most beneficial in *planned* FonF activities. Support for this claim can be found in studies by Doughty and Varela (1998), McDonough and Mackey (2006), and Han (2002). In these studies, recasts were found to be effective when frequently and consistently provided in response to a single erroneous form. Teachers in Han's longitudinal study of the effects of recasts on tense use were careful to provide consistent recasts in response to learners' inconsistent tense use; if a learner chose to use the past tense the researcher focused on consistent use of the past tense in providing recasts. As well, the researcher was careful not to recast a correct

switch of tense. Learners were found to have benefited from recasts in this context, as measured by the same learners' performance on a pre-test and two subsequent post-tests.

To summarise this section, recent research into interactional feedback has largely focused on recasts and whether they are as effective as prompts and elicitations in promoting L2 development. While recasts and elicitations are both implicit types of negative feedback, recasts may be more ambiguous in negotiations for meaning. However, the research is inconclusive when it comes to measuring feedback effects. The following section discusses these mixed results and examines *why* the results are so inconsistent.

2.3.2 *Mixed Results for Feedback Effects*

The effectiveness of feedback has been measured in a variety of ways including learner uptake (e.g., Braidı, 2002; Lyster & Ranta, 1997; Panova & Lyster, 2002), the spontaneous production of a higher developmental form (e.g., Mackey & Oliver, 2002; Mackey & Philp, 1998; McDonough & Mackey, 2006 and Spada & Lightbown, 1999), stimulated recall interviews (Mackey *et al.*, 2007), and tailor-made individualised tests designed to elicit the targeted form (Nabei & Swain, 2001). As noted by Ellis (2001), different measures of acquisition may lead to different results. The mixed results observed in the recent literature on recasts alone attests to this. This section provides a comparison of a few studies of the effects of recasts to illustrate how different measures of acquisition provide a confusing picture of their role in the L2 classroom. In addition, this section will discuss other factors that render the results of many feedback studies

incongruous with others, including developmental readiness, the method used for eliciting target forms, and operationalisations of variables.

Studies by Mackey (1999), Mackey and Philp (1998), and Lightbown and Spada (1999) used question forms as the measure of development. Examples of question forms and developmental stages are provided on page 52. Question forms can be easily elicited in a number of tasks (i.e. picture difference, picture story, etc.) and empirical research for the developmental stages of question formation is quite strong (Mackey & Philp, 1998; Pienemann, Johnston, & Brindley, 1988). L2 learners of English progress through stages of acquisition which are similar to those of children learning question development in English as a first language. As such, L2 development can be measured as the number of stages through which a learner progresses in the developmental sequence, or as the increased amount of grammatical questions produced within a single stage. While Pienemann and his colleagues (Pienemann, Johnston, & Brindley, 1988) suggested that two different productions of two different structures is enough to indicate the acquisition of a given stage, others have applied a more rigorous requirement of two different structures produced in two different tasks (e.g., Mackey, 1999; Mackey & Oliver, 2002; Mackey & Philp, 1998); this is likely a better indication of whether a given stage has been reached. For example, an interesting study by Mackey and Oliver (2002) provides evidence that effects of recasts are delayed; learners receiving recasts actually performed significantly better on delayed post-tests than on immediate post-tests.

Closely related to the concept of measuring the effects of feedback according to developmental stages is the issue of developmental readiness. It is important to note that interactional feedback is a means of providing FonF in negotiations for meaning. As

such, it is a method of “promoting the processes involved in natural language acquisition,” not of changing them (Ellis, 2001, p. 4). In a study by Mackey and Philp (1998), learners were judged to be either “ready” or “unready” to advance to higher stages of development. Learners’ scores on a proficiency assessment conducted by the participants’ school determined whether they were “ready” to acquire higher stage question structures. Transcripts of a conversation task performed by the “unready” participants were also examined by the researchers to confirm their developmental level. Their study provides evidence that beginner learners may be “unready” to produce questions at higher stages of development (Mackey & Philp, 1998). Specifically, learners who were developmentally ready were able to produce higher stage questions after intensive recasts, while those learners who were unready did not show such benefits of recasts. This study allows us an important insight into the contexts where recasts might be most effective. First, it is possible that studies claiming recasts to be ineffective were targeting L2 forms that were beyond the learners’ level of readiness. Second, it is possible that recasts may be more effective with more advanced learners; the “readies” were already producing higher stage questions than the “unreadies” at the beginning of the study, so it is possible that more advanced learners are better able to notice and respond to recasts as feedback than learners who are just beginning.

In contrast to the use of developmental stages, Braidí’s (2002) study of the role of recasts in interactions between native speakers and non-native speakers examined only the incorporation of recasts in the immediate turn. Braidí’s study aimed to confirm the existence and the role of negative feedback and of recasts in a non-classroom setting. Negotiations were classified as one-signal or extended interactions; meaning was

negotiated with only one signal of miscomprehension in the former, while multiple signals were typical of the latter. Braidı predicted that the length and difficulty of the negotiations would affect the type of feedback provided by native speakers. Ten adult native English speakers and ten adult Japanese speakers were randomly assigned to ten NS-NNS dyads, and each dyad completed four communicative tasks within one hour. A native speaker's response was coded as a recast if it included the content words and modified or added to the learner's previous utterance in some way. The results indicate that recasts comprise 9.56% of the native speakers' responses, and occurred in both one-signal and extended negotiations, although a larger number of recasts occurred in extended negotiations than in one-signal negotiations. Braidı suggests that as interaction becomes more difficult, recasts increase in frequency. Thus, recasts as negative feedback are available to learners when they are likely to use them. However, while the purpose of the study was to investigate the role of recasts in interaction, no generalisations can be made regarding their role in facilitating L2 acquisition as there is no measure of later incorporation or of the use of recasts in general.

In a similar study of learner uptake following different types of interactional feedback, Lyster and Ranta (1997) distinguish between learner uptake and learner repair, where uptake may include or result in repair. For the purposes of their study, learner uptake is defined as a learner's utterance that immediately follows feedback from the teacher and is in some way a response to the teacher's attempt to draw the learner's attention to a particular aspect of the incorrect utterance. As such, learner uptake consists of "repair" and "needs repair": a learner's utterance in response to a feedback move either consists of a correct reformulation of the incorrect utterance (repair) or remains incorrect

(needs repair). In their study of French immersion classes, Lyster and Ranta found that recasts only resulted in learner uptake 30% of the time, while elicitations and metalinguistic feedback were more successful.

Similarly, Lyster (2004) found that FonF was more effective when combined with prompts than with recasts or no feedback in a study of fifth-grade learners of grammatical gender in French. Lyster suggests this may be due to the fact that recasts are somewhat ambiguous in nature; by providing learners with “correct target forms, which frequently co-occur with signs of approval” (Lyster, 2004, p. 404). In his study of four francophone teachers of eight fifth-grade French immersion students, three groups received form-focused instruction (FFI) while the fourth group continued with its regular curriculum without any special instruction. Two of the FFI groups received feedback (one group received prompts, the other recasts). The study followed a pre-test, post-test, and delayed post-test design, with the post-test given after an eight week period. While the FFI-prompt group consistently out-performed the FFI-recast group, the FFI-recast group also significantly outperformed the comparison group. However, there were no significant differences between the FFI-recast group and the FFI-only group on the second post-test. Both recasts of ungrammatical utterances and repetitions of perfectly acceptable and grammatical utterances seem to be interpreted as confirming or disconfirming the meaning of a learner’s utterance, instead of the form. This suggestion highlights the problem of *how* recasts should be provided to learners in SLA research. It may be that some tasks or contexts that focus on meaning rather than form are less likely to induce students to recognise a recast in response to an error of form.

It is possible that the method used to stimulate negotiation for meaning, as well as learners' uptake and repair in response to teachers' feedback, may influence learners' responses to recasts and other forms of interactional feedback. Braidí (2002) notes that conversations may be less goal-oriented than task-based interactions and interlocutors therefore may not feel the need to resolve a communication difficulty in a conversation. While the term "task-based" has often been used in place of "communicative," for the purposes of this thesis, it is important to distinguish between an "activity" that could refer to controlled or more open classroom procedures and a "task" (and by extension, task-based interaction) which is geared to eliciting practice of a linguistic form presented earlier (Crookes & Chaudron, 2001).

Research on interactional feedback has been carried out with different methodologies in a variety of settings and has provided important pieces to the puzzle of feedback effects. However, regardless of whether uptake or developmental stages are used to measure acquisition, or whether the research takes place in an L2 classroom or between researcher-student dyads, it is important to isolate the effects of each type of feedback as much as possible. In a study of learner uptake of different types of feedback in communicative classrooms across four instructional settings, Sheen (2004) found that recasts are highly variable in number and across pedagogical methods and settings. It was found that recasts are less frequent than other types of feedback in Canadian immersion and ESL settings than in New Zealand ESL and Korean EFL settings. It is suggested this may be due to a greater amount of formal instruction in the New Zealand and Korean settings than in the Canadian classrooms. However, it is difficult to interpret these findings as a direct reflection of the amount of formal instruction received as there

are too many other possible influences to isolate the source of any differences between instructional settings. For example, the four studies were highly varied in terms of the amount of L2 instruction; French immersion students in Canada spent about 60% of their school day in French instruction, including subject-matter classes such as math, science and social studies. In contrast, learners in an intensive ESL classroom in New Zealand received only four hours of instruction a day, and EFL learners in Korea attended classes for about six hours each week. The ESL classrooms in Canada and New Zealand were quite different in terms of their L1 and English proficiency: the learners in the Canadian ESL classroom shared the same L1 and were just beginning their English education, while the learners in the New Zealand setting were from different L1 backgrounds, and were at an intermediate level of proficiency. These are only a few examples of the diversity between the four settings discussed by Sheen; thus, any differences observed in the frequency and effectiveness of recasts can not be directly related to any specific factor because recasts as a type of feedback were not isolated from myriad other factors.

A final problem that makes it difficult to interpret the results concerning interactional feedback concerns the use of varying operationalisations of different types of feedback. For example, recasts were defined by Mackey and Oliver (2002) as “more target-like alternatives which follow a learner’s non-target-like utterance (p. 464)” in a study of the effects of interactional feedback on L2 development in children. One group received interaction with feedback while the other group did not receive any feedback. While the results seem to provide support for beneficial effects of recasts, the operationalisations of “feedback” and “negotiation” make the results difficult to interpret. In their study, negotiation is seen as a type of feedback that may or may not coincide with

recasts. This definition is quite different from Long & Robinson's description of feedback, where negotiation "elicits negative feedback, including recasts" (Long & Robinson, 1998, p. 23). Consequently, it is difficult to compare the results of Mackey and Oliver's study to research that has investigated recasts as feedback that is often provided within negotiation (e.g., Braidi, 2002; Leeman, 2003; Loewen & Philp, 2006). Thus, including negotiation as a type of feedback instead of a situation where feedback can occur is somewhat inconsistent with previous research and renders the results unclear.

Similarly, Lyster (2004) used the term "prompts" to refer to four types of teacher response to errors, including clarification requests, repetitions, metalinguistic clues, and elicitations. Lyster claimed that each of these moves differ from recasts in one important way: they withhold the correct form in order to provide learners an opportunity to self-repair. However, Lyster's results that prompts are more effective at drawing learners' attention to form than recasts are hardly surprising when, in essence, the effects of four feedback types were compared to one. In addition, these results are not easily comparable to those of other studies that have examined each of these "prompts" as a single type of feedback (e.g., Lyster & Ranta, 1997; Sheen, 2004). In consequence, it is difficult to interpret exactly *which kind* of prompt is potentially better than recasts in conversational interaction.

In summary of this section, it is difficult to make generalisations about the effects of recasts on the basis of these combined studies; while one measures only the uptake produced in learners' private speech, another does not take into consideration the possible effects of recasts beyond the immediate turn, while still another examines the effects of recasts as positive, instead of negative evidence. In addition to the conflicting results and

difficulties inherent in pinpointing the source of any beneficial effects of recasts, there are several challenges involved in measuring such effects. These challenges are explored in the following section.

2.3.3 The Challenge of Investigating Feedback

Inherent in any investigation of feedback effects is the problem of how to measure such effects. The L2 classroom is a rather unpredictable and challenging environment in which to carry out experimental studies that seek to isolate certain variables. This section discusses how previous research has addressed these issues.

Because of the mixed results of research into the effects of recasts and other types of interactional feedback, there have recently been some efforts made to examine specific aspects of recasts and the contexts in which they might be most beneficial. For example, Leeman (2003) suggests that the juxtaposition of recasts and learners' original utterances may emphasize any target form that is found in the reformulation but is absent in the original utterance. In a study investigating the source of effects of recasts, Leeman explores implicit negative evidence and salience-enhanced positive evidence in recasts. Four types of interactional input were isolated to examine whether recasts are beneficial in developing noun-adjective agreement in Spanish, and whether benefits can be attributed to either negative evidence or enhanced salience of positive evidence. All of the 74 participants were L1 speakers of English in first-year Spanish courses randomly assigned to a recast, negative evidence, enhanced salience, or control group. In the recast group, the researcher's responses provided target-like reformulations of learners' utterances, thereby providing implicit negative evidence about the unacceptability of

utterances while enhancing the salience of positive evidence. Participants in the negative evidence group received implicit negative feedback only, while the enhanced salience and control groups received no feedback.

The results of participants' post-test scores suggest that only the recast and enhanced salience groups engaged in long-term development. Thus, Leeman suggests that recasts are successful insofar as they provide enhanced salience of positive evidence. However, Leeman fails to discuss the fact that recasts are often mistaken by learners as clarification requests. In her study, information-gap activities were completed between native-speaker - non-native-speaker dyads and were designed to fit within a meaning-based approach to language learning. Tasks included completing a picture, selecting items from a catalogue, and placing items in the correct location. Consequently, any differences found between groups who received positive or negative input may be a result of the learners understanding the input as a confirmation or a negation of the students' request for information. In such cases, it is debatable that any such enhanced salience exists; at the very least, any increase in saliency is somewhat masked and is therefore questionable. While the tasks were successful at eliciting learners' utterances and subsequent teacher recasts, an inherent problem with researching recasts within meaning-based contexts is reflected in this study. On the other hand, it is interesting that the enhanced salience led to greater learner accuracy in a meaning-based interaction. If the advantages of the enhanced salience group can be attributed to increased learner attention to target forms, then it may be possible to increase learner attention to form without sacrificing a focus on meaning. Nevertheless, further research is needed to establish the advantages of enhanced saliency of positive evidence.

Whether the result of positive or negative evidence, a review of the literature on recasts by Nicholas, Lightbown and Spada (2001) concluded that recasts are most effective when learners realise the recast is in response to the accuracy (or inaccuracy) of the form, not the content, of their original utterance. Their conclusion suggests that the conflicting results concerning interactional feedback may be due to its implicit nature; that is, interactional feedback may prove to be quite effective in those cases where the students actually perceive the teacher's response as feedback to a problematic form. This raises two significant problems: (1) how can researchers determine how students interpret interactional feedback? And (2) how can researchers ascertain whether feedback has positive effects for learners who do *not* perceive or immediately respond to the feedback during the interaction? Mackey, Gass, and McDonough (2000) investigated the first of these issues and found that learners had the most difficulty in recognising feedback concerning morphosyntactic errors, although they were relatively successful in accurately recognising feedback in response to semantic, lexical and phonological errors. In their study, 17 L2 learners (10 learners of English as a second language and 7 learners of Italian as a foreign language) were videotaped participating in task-based dyadic interaction with native speakers. Immediately following the task, the learners participated in a stimulated recall task; learners watched the videotapes of their interactions and were asked to pause the tape whenever they wished to describe their thoughts at any given time during the interaction. It is interesting to note that feedback in response to morphosyntactic errors (the least likely to be noticed) was most frequently provided in the form of a recast; thus, it is possible that the reason the feedback to morphosyntactic errors largely went unnoticed was because the learners did not perceive

the recast as feedback. However, this study did not empirically measure the effectiveness of feedback, and thus should not be taken as evidence that recasts and other types of implicit feedback are not effective.

Also relevant to this first issue of how students interpret interactional feedback is Havranek's explorative study of situational, linguistic and personal factors that might promote or impede the effectiveness of corrective feedback (Havranek, 2002). In her study, Havranek observed 207 EFL learners at various proficiency levels from 10-year old elementary beginners to adult university students studying English. All learners were native or near-native speakers of German. Four levels of English were observed at the school, and two levels were observed at the university. After an observation period of 2 weeks for the school and 5-8 weeks at the university, a little over 50% of the learners who received feedback used the same structures correctly on a subsequent post-test (it is not mentioned *when* the post-tests were administered). The results showed that recasts were not as effective as explicit elicited self-correction, where the teachers indicated there was an error but did not provide the correct form. On the other hand, all forms of corrective feedback were found to be most effective when the learners provided the correct forms immediately following the feedback. Surprisingly, the student peers who heard the correction of a fellow student performed *better* on the post-test than the students who were corrected; on the post-test they used the correct forms of the structures that their peers had used incorrectly during the observation period. Unfortunately, this study did not include a pre-test, so it is impossible to know whether the students who heard the correction already knew the forms in question and were thus able to use the forms correctly on the post-test. Nevertheless, this study provides evidence that learners

may not be required to participate in the interaction to benefit from the corrective feedback provided during the exchange.

The findings of Havranek's study are supported by Ohta's (2000) study in which the definition of uptake is extended to include individual responses from learners who are not directly addressed with the feedback. These responses are neither directed at the teacher nor at the student who received the feedback, but are typically private responses (hereafter referred to as private speech) that the learners utter in response to the L2 input that is available around them in the classroom. The concept of private speech comes from a sociocognitive perspective where language development and production are "mutually constitutive" (Ohta, 2000, p. 52). While traditionally a source of data for first language researchers, private speech research contributes to L2 research insofar as students' internalization and L2 problem-solving can be observed as they think aloud. In her study of 7 Japanese language learners' self-addressed utterances, the effects of recasts are examined in terms of learners' private responses to corrective feedback directed at other learners. Her findings suggest that learners are "mentally active in attending to and analyzing recasts (p. 49)." The data includes transcripts of the private speech of one second-year and three first-year university students collected via individual microphones. The first-year Japanese textbook incorporated a grammatical syllabus overlaid by a topical one, while the second-year textbook was more focused on grammar. Characteristics of private speech include reduced volume, lack of adaptation to an interlocutor, and lack of response from an interlocutor. Learners were considered conversational participants not only when they were the addressees of corrective feedback, but also when they were auditors of such feedback. It was found that learners

produced private speech most often when they were auditors, not addressees. Results show that when learners produce ill-formed utterances in their private turns, they may contrast with utterances of others which function as incidental recasts. Examples include repetition of recasts directed at another student, individual response within a choral context and subsequent modification based on the teachers' succeeding utterance, perceived recasts functioning as either elicitations or prompts for other students, and using recasts as confirmations of self-correction. Thus, private speech provides insight into the mental activity in which learners engage concerning corrective feedback. It may be that what learners produce in private speech forecasts what they will later produce in social speech; private speech is likely an indicator of developmental processes.

An important implication of these studies, then, is that any study intent on evaluating feedback effects should not investigate the negotiation of form alone. In other words, feedback effects may not be immediately observable or limited to a learner's language production within a given interaction. The absence of a learner's immediate uptake or repair within a negotiation of form does not necessarily entail that the learner has not processed the feedback (Mackey & Philp, 1998; Nabei and Swain, 2001). It would be extremely difficult to show that a corrective move by a teacher results in complete understanding and repair of the error by the learner. Such a claim would have to be supported by "long-term observation of the learner's production of the corrected structure while at the same time making sure that there is no further input of the same structure, ruling out any other source of learning" (Havranek, 2002, p. 256).

As mentioned above (p. 27), a second problem with measuring feedback effects concerns a learner's immediate response (or lack of response) to feedback. Specifically,

it is difficult to determine whether a learner's immediate response to feedback, or "uptake," (Lyster & Ranta, 1997) is an accurate measure of language learning (Braid, 2002). Conversely, does the lack of learner uptake indicate that no learning or internalisation of the feedback has occurred? (Ohta, 2000). McDonough and Mackey (2006) explore these questions in their study into the relationship between recasts, learners' responses to recasts, and their development of ESL question forms. In a 9-week pre-test/post-test design, 58 Thai EFL learners were analyzed for the occurrence of recasts and their responses to recasts that targeted developmentally advanced question forms. The study involved a treatment condition where learners received recasts and a control group where ungrammatical utterances were ignored. The results showed recasts to be a significant predictor of ESL question development. However, immediate repetition of the recasts were *not* associated with question development; that is, students were able to produce higher stage questions on subsequent post-tests even in those instances when they did not immediately repeat the recast. In their study, only 2 of the 19 learners in the no-feedback group advanced to a higher developmental stage on the post-test, while 23 out of 39 learners who received feedback in the form of recasts advanced to a higher stage. This study by McDonough and Mackey provides evidence that recasts may be effective even if learners do not repair their errors in immediate uptake.

The studies discussed above have all made important contributions to our understanding of interactional feedback and how it might benefit L2 learners. Specifically, research has largely focused on the efficacy of recasts and has attempted to determine whether the high frequency of recasts in various L2 contexts is an accurate

reflection of their usefulness. In particular, there is some evidence that recasts are effective when provided consistently (Han, 2002) and when learners are developmentally ready to attend to them (Mackey & Philp, 1998). Additionally, both recasts and other forms of corrective feedback are likely less ambiguous when provided in form-focused rather than meaning-focused negotiations where, for example, feedback focuses on a single linguistic item (Doughty & Varela, 1998; Lyster, 2002). However, they have also highlighted some of the difficulties involved in measuring the effects of feedback. While some researchers have used immediate uptake and repair to measure feedback effects, others have used learners' use of higher developmental stages. In addition, there is a need for more longitudinal studies that would allow researchers a better understanding of how feedback affects learners over time. Because it is difficult to compare or make sense of these results, there is also a need for future research to carry out studies that reflect both experimental and observational qualities. The following section explores these issues.

2.3.4 Research Needed

As mentioned at the end of Section 2.3.3, it can be difficult to interpret the findings of research concerning the effects of interactional feedback. However, it is important to be able to compare results across studies in order to confirm or negate the benefits of a given type of feedback. This section focuses on two main issues: (a) the need for longitudinal studies to be carried out in order to observe long-term effects of feedback and (b) the need for “hybrid” research (Ellis, 2001) and quasi-experimental studies that include both quantitative and qualitative procedures. Such studies enable the

researcher to investigate more than learners' immediate responses to feedback while still incorporating empirical and statistical methods of data collection and analysis.

First, in order to judge whether the effects observed in learner uptake are truly representative of language learning, it is necessary for studies to be more longitudinal in design. As mentioned by Ellis (2001), surprisingly few studies of classroom learners have been longitudinal. Surely, in order to establish the effectiveness of corrective feedback any observed effect must be sustained over time (Lightbown, 1998). In addition, there are very few studies examining delayed effects of feedback. Two examples of longitudinal studies are those of Mackey (1999) and Mackey and Oliver (2002) where both sustained and delayed effects of recasts were investigated. Mackey and Oliver found that significantly more child ESL learners in dyads with adult native-speakers of English showed sustained development in the dyads that received recasts than in the dyads that did not. Even more suggestive is the fact that sustained development was not only maintained but also *improved* over time; the post-tests were given immediately following treatment, one week after treatment, and again after three weeks. Such long-term effects may not have been revealed if the length of the study had been shortened. On the other hand, it can be difficult to collect adequate data in a classroom environment over a long duration without intruding on the regular course syllabi. While it is easier to conduct longer studies outside of the L2 classroom, it would also be easier to assess pedagogical benefits of feedback if those benefits actually took place during a typical L2 class.

As suggested above, a second problem yet to be addressed in the literature on interactional feedback concerns the rather large gap between observational and

experimental research. While there has been some interesting observations made in L2 classrooms (e.g., Ellis *et al.*, 2001; Lyster & Ranta, 1997; Panova & Lyster, 2002), many of these studies have focused on incidental FonF, making it difficult to control for different types of feedback provided in a regular L2 classroom setting. Conversely, experimental studies have largely consisted of researcher/student or native speaker/non-native speaker dyads that take place outside the classroom (e.g., Braid, 2002; Mackey & Oliver, 2002) and are not representative of typical language learning environments.

One of the inherent difficulties in conducting experimental research in SLA is the spontaneous nature of the L2 classroom; teachers may have a hard time restricting feedback to a single type, and it is certainly difficult for a single teacher to provide interactive feedback to each student in response to every error. Basturkmen, Loewen, and Ellis (2004) found a discrepancy between teachers' stated pedagogical beliefs and their pedagogical practices in the L2 classroom. Their finding illustrates that controlling the implementation of treatments carried out by participating teachers can be quite difficult. This is an ongoing problem in the field of SLA; while more control is attainable in laboratory-based experiments, the practical benefits of classroom-based research can not be denied. These problems are reflected in the relative dichotomy between observational and experimental research on interactional feedback. Thus, the challenge presented to researchers interested in quasi-experimental studies of feedback in L2 classrooms is how to control for feedback type without compromising the integrity of the L2 classroom.

One method of carrying out such research is the use of hybrid studies that allow for certain aspects of the class to be controlled by the researcher for experimental

purposes while collection of the resulting data is done as unobtrusively as possible. One example of such hybrid research is Williams' (1999) study of classroom learners at four different levels of language proficiency. In her investigation of learner-generated attention to form, Williams found that the percentage of language-related-episodes (that is, dialogue that focuses on a linguistic form instead of meaning) was clearly higher with more proficient learners. This study explored learners' interactions in a real ESL classroom and employed statistical methods of data analysis on data that was qualitatively collected through classroom recordings.

An example of quasi-experimental research carried out in an L2 classroom is Ammar and Spada's (2006) study of recasts and prompts and their effects on learners of different proficiency levels. Learners in this study were all students of three intact intensive ESL classes and were assigned to one of three conditions: one class received recasts, one received prompts, and the control group did not receive either form of interactional feedback. The target form was third-person possessive determiners *his* and *her*. The study followed a pre-test, immediate post-test and delayed post-test design. Students in both of the feedback groups outperformed those in the control group. The results also suggest that prompts were more effective than recasts, but that learners of higher proficiency levels were more likely to benefit from recasts than learners of lower proficiency. In addition to providing some interesting results concerning the role of developmental readiness and feedback effects, this study also makes an important contribution to the literature as a model of how quasi-experimental research can be accomplished in the L2 classroom.

With closer inspection of these and other studies investigating recasts and other types of interactional feedback, it becomes clear why the research to date has produced such mixed results. Differences in measures of acquisition, operationalisations of feedback types and research methodologies render interesting results that are nevertheless difficult to interpret as a single body of research. In addition, the spontaneous nature of the L2 classroom represents a challenge for second language acquisition (SLA) researchers in that it provides a difficult environment in which to restrict interactional feedback to a single type. Recasts, in particular, have been under the microscope of many researchers and have been frequently compared with other types of interactional feedback, such as elicitations and other prompts (e.g., Doughty & Varela, 1998; Lyster & Ranta, 1997; Nassaji, 2007a; Sheen, 2004). In order to address these issues, the goal of this study is to examine the effects of different types of interactional feedback, specifically recasts, elicitations, and recasts plus elicitations on the development of question formation by ESL learners in a previously established and typical classroom environment. Question formation was chosen as the target structure for this study because learning can be measured with relative accuracy by the developmental stages identified by Pienemann, Johnston and Brindley (1988). As well, the stage of questions learners produce throughout the seven weeks of this study should accurately indicate whether acquisition has occurred; this study will allow for the possibility of delayed effects of feedback in a way that immediate uptake in a short-term study might not. Although studies have provided evidence for the importance of feedback provided in conversational interaction (as mentioned above), the question of which types of feedback are the most effective remains open. Although teachers tend to use recasts more than any

other type of corrective feedback (Lyster & Ranta, 1997; Sheen, 2004), other types may prove to be more effective. Thus, by conducting quasi-experimental research in typical L2 classrooms, this study intends to contribute to the debate concerning interactional feedback and recasts, in particular.

Chapter Three

METHODOLOGY

This research followed a longitudinal quasi-experimental design in order to test for the effects of different types of interactional feedback. The research schedule involved a pre-test, an immediate post-test, and a delayed post-test. This study was incorporated into the participants' regular course content and focused on the teachers' responses to ungrammatical utterances with a single type of interactional feedback. This chapter provides a detailed description of the procedures used to collect and analyse the data. The first section presents the background information of the participants and the recruitment process. The second section outlines in detail the data collection procedures, research schedule and design, including the materials used for each of the test and treatment tasks. The third section summarizes the process of categorising the questions into the various stages of question development and coding the data. The fourth and final section describes the statistical analyses of the data; both qualitative and quantitative analyses were done to examine the frequencies and means of question production across the four groups, as well as the results of One-Way and Repeated Measures ANOVAs to compare these means.

3.1 Participants

Teacher recruitment began at the beginning of the fall semester of 2006, at which time I presented the research goals of the study to the staff at an English language institution in B.C., Canada. Two teachers and 63 students participated in this study.

Both of the teachers were female native speakers of Canadian English and had considerable teaching experience in both ESL and EFL contexts prior to the time of this study. Specifically, one teacher had been teaching for ten years, including four years of experience in Korea and five years teaching at the school where the data were collected for this study. The second teacher had also taught in several EFL contexts before teaching at the school in question for five years before the study began. The teachers volunteered to designate a significant amount of their class time to the study.

All of the student participants (44 females and 19 males) were registered in regular classes at the same secondary education institution when this study began (information regarding the student background can be found on page 41). Student recruitment involved a presentation to each of the four participating classes. In each presentation, I explained that the focus of the research was interaction between teachers and students, and emphasized that their participation in the study was completely voluntary. The students were not required to spend any time outside of their regular class hours, and the research tasks would only use vocabulary from their regular course materials. The teachers were present during these presentations and repeatedly explained to the students that their performance during the research tasks would not have any influence on their regular course grades or evaluations.

No beginner-level students participated in this study because it is likely that beginner learners would produce a high number of Stage 1 questions consisting of only a single word with a rising intonation. These questions are hard to correct with implicit feedback because the meaning of the questions is often unclear. For example, the utterance “*Deer?*” could mean either, “*Is this a deer?*” or “*What is a deer?*” or possibly

“*Do you like deer?*” Thus, it is more difficult to implicitly provide meaningful and appropriate feedback to beginner students than intermediate or advanced students.

Conversely, advanced learners may not have made enough errors for a sufficient analysis of feedback effects; thus, intermediate learners were determined to be the most appropriate participants for this research.

The proficiency level of all participating students had been measured by administrators at the school before recruitment for this study began. All of the students were placed at an intermediate level (level 3 of 5 possible levels); that is, students were able to satisfy most survival needs and demands, formulate questions when necessary, and were proficient enough to understand simple conversations. Placement of students into appropriate classes is a two-step process. First, the Secondary Level English Proficiency (SLEP)⁵ placement test is administered to all students and measures their ability to understand spoken and written English (See Appendix A for an example of the SLEP test). Second, the teachers of each class make close observations of students’ spoken and written performance. The first week of classes are typically spent evaluating students’ proficiency during a variety of communicative tasks. At the end of the first week, any students who are judged to be at a higher or lower level of proficiency than the level of the class are moved to a more appropriate level. The placement procedures conducted by the administrators were considered to be sufficient for the purposes of this research.

While the extent of previous English education was highly varied, all students had been studying English for a minimum of two years. There was also a wide range in

⁵ The SLEP Test is a creation of the Educational Testing Service (ETS). ETS is also the creator of the *Test of English as a Foreign Language*TM (TOEFL).

length of stay in Canada; while some students were only attending one three-month session, other students intended to enrol in regular courses at a university once they had achieved the proficiency requirements of the university in question. All of the students were studying English for either business or academic purposes (either in their country of origin or in Canada). The themes, activities, and structure of every course at the school are determined by the teachers. At the beginning of term, teachers select specific outcomes from a range of level-appropriate items outlined by the school in a teacher's handbook. While the vocabulary and activities vary with each teacher, all classes are directed towards the particular goals outlined for each level. Some of the speaking goals of the Level 3 classes include being able to respond to questions in predictable contexts, to develop content words and idiomatic language, and to clarify meaning by rewording or rephrasing utterances. The vocabulary for two of the participating classes included animals, Halloween and crime. The vocabulary for the other two classes included relationships, dating and idiomatic expressions for feelings and daily life.

Information regarding the students' gender, age, language and educational background, previous time spent in Canada and length of stay was collected from a participant background questionnaire at the beginning of the study (see Appendix D). Students came from a variety of linguistic backgrounds; first languages included Arabic, Chinese, French, Japanese, Korean, Spanish and Turkish. Table 3-1 provides a summary of the first language, gender, age and educational backgrounds of the students. While all of the 63 students attended most of the treatment sessions, 10 students were excluded from the data analyses because they were absent for at least one of the pre- or post-test

sessions. The data presented in Table 3-1 include only those 53 students who attended every session, including pre-test, treatment, and post-test tasks.

Table 3-1

Participant Summary

First Language	No. of Students	Gender	Age	Years of English Studies
Arabic	3	3 male	19-22	8-13
Chinese	12	5 male 7 female	18-33	3-14
French	1	1 female	over 50	several
Korean	11	4 male 7 female	19-25	3-15
Japanese	19	3 male 16 female	18-21	6-14
Spanish	6	6 female	17-26	3-17
Turkish	1	1 male	19	2
Total/(Mean)	53	16 male 37 female	22*	(9)

* *This excludes the age of one participant whose exact age is unknown.*

3.2 Research Design

This section presents the research schedule and design, the data collection procedures, and describes the tasks and methods involved in the following four conditions: control, recast, elicitation and recast + prompt.

Each class was randomly assigned to one of the four conditions. As mentioned above (page 41), the content of lessons outside of the experimental research sessions varied among the four classes. As some of the general outcomes of this intermediate level course included students' ability to respond to questions and ask for advice, some instruction and class activities involved question formation. However, the main foci of these classes were not question formation but theme-based vocabulary and the introduction of grammatical structures such as present, past, and future clauses, the present perfect, and the use of gerunds and adjective clauses (Instructors' Handbook, 2006). During the regular classes that were not included in this study, teachers corrected any type of student error as they saw fit. As regards the research sessions, each of the treatment groups (recast, elicitation and recast + prompt) received a single type of feedback once a week for four weeks. Participants in the elicitation group received elicitation as the sole type of feedback (see Example 1), participants in the recast group only received corrective reformulations of their utterances (see Example 2) and participants in the recast + prompt group received both recasts and elicitation in the form of prompts (see Example 3). All of the following examples were taken from the research data.

(1) Elicitation

Student: Where is the grey whale live?

Teacher: What's the question?

In Example 1, the teacher responds to the student's question using an elicitation without *explicitly* stating that an error has occurred or where the error may be in the utterance. By asking the student to reaffirm the question, the teacher is *implicitly* indicating that something in the student's utterance is not acceptable. The teacher's response could be interpreted as either an indication that the utterance was incorrect or as a conversational turn, such as a clarification request.

(2) Recast

Student: Where is the grey whale live?

Teacher: Where does the grey whale live?

In Example 2, the teacher provides a recast of the student's utterance without altering the meaning of the question. As with the previous example, the teacher's feedback is implicit; the student may either interpret the reformulation as corrective feedback or as a comprehension check in conversation.

(3) Recast + Prompt

Student: Where is the grey whale live?

Teacher: What's the question? Where does the grey whale live?

In this example, the teacher's response includes both of the elements illustrated in Examples 1 and 2. In this case, the feedback implicitly indicates that there is something

unacceptable about the student's utterance *and* provides a grammatical reformulation (recast). As with the recast and elicitation groups, the teacher's response could be interpreted as a conversational turn instead of implicit corrective feedback.

Throughout the entire duration of the study (seven weeks), the teacher of the control group did not provide any corrective feedback in response to any errors that occurred during the question elicitation tasks that were observed for this research. The typical response to both grammatical and ungrammatical questions was a verbal "Okay" or a non-verbal gesture such as a nod. Similarly, no corrective feedback was provided to learners in any of the four groups during the pre-test and post-test tasks.

Students in all four groups completed a pre-test task one week before the first treatment began. The students' performance on the pre-test task determined participants' level of development in question formation before treatments. Students also completed two post-test tasks after the final treatment session. The level of questions students produced on the pre-test task were compared to those produced on an immediate post-test task the day after the final treatment and on a delayed post-test task two weeks after the final treatment. The pre-test, treatments, and both of the post-tests were in the form of question elicitation tasks. Table 3-2 displays the schedule for the pre-test, treatments, and post-tests.

Table 3-2

Research Schedule and Design

Week	Day	Group			
		Control	Elicitation	Recast	Recast + Prompt
1	1	Pre-test	Pre-test	Pre-test	Pre-test
		Tic Tac Toe	Tic Tac Toe	Tic Tac Toe	Tic Tac Toe
2	1	Treatment	Treatment	Treatment	Treatment
		Crossword 1	Crossword 1	Crossword 1	Crossword 1
3	1	Treatment	Treatment	Treatment	Treatment
		Story Completion 1	Story Completion 1	Story Completion 1	Story Completion 1
4	1	Treatment	Treatment	Treatment	Treatment
		Crossword 2	Crossword 2	Crossword 2	Crossword 2
5	1	Treatment	Treatment	Treatment	Treatment
		Story Completion 2	Story Completion 2	Story Completion 2	Story Completion 2
	2	Post-test 1	Post-test 1	Post-test 1	Post-test 1
		Tic Tac Toe	Tic Tac Toe	Tic Tac Toe	Tic Tac Toe
7	1	Post-test 2	Post-test 2	Post-test 2	Post-test 2
		Crossword 1	Crossword 1	Crossword 1	Crossword 1

3.2.1 Data Collection Procedures

I recorded each class with a digital video-camera. The video-camera was positioned near the front of the class for the best audio and visual quality; however, I was careful to place the video-recorder off to the side of the classrooms to reduce any anxiety the students might have felt in front of the camera. In addition, I remained at the side of

the class for the duration of the sessions so as not to distract the students or to interfere with the tasks. I also gave instructions about each task at the beginning of every session to ensure that the information was given consistently to all groups. I was present throughout all of the sessions as a non-participant observer and took notes on the level of questions students produced and their responses to a given type of feedback.

3.2.2 *Pre-test*

A modified version of Tic Tac Toe was used for the pre-test and the immediate post-test.⁶ The original version of the game was adapted so that two teams of students could play at once, instead of only two individual students. Each class was divided into two teams and students in each team were given blue or yellow stickers to identify their team. The students were presented with a game board consisting of nine vocabulary cards placed face-down in three rows of three. Each card consisted of one vocabulary word or phrase that the students had already covered during previous classes. The teachers chose the vocabulary to be used for each session (including treatment and post-test sessions) so that the students were always presented with words and phrases they had practiced in class. The vocabulary varied among classes as the themes of the vocabulary cards depended on each teacher's previously planned course syllabus. Students took turns selecting a card from the grid and providing questions that included the vocabulary word or phrase on each card. Students provided a single question per turn, with each student taking at least two turns. Each time a student completed a turn (by producing a question that used the vocabulary word on the selected card) the student placed one of

⁶ Tic Tac Toe normally consists of a blank grid that two players fill in with X's and O's to make complete rows in vertical, horizontal and diagonal directions. The game is over when no squares in the grid remain and the player with the highest number of completed rows wins the game.

their team's coloured stickers over the corresponding square on the board. Once a team had made a complete row with their stickers, the teachers reorganised the pictures and began again. A team won the game by making the greatest number of rows with their coloured stickers. During this pre-test task, the teachers did not provide any corrective feedback to the students in response to ungrammatical utterances.

3.2.3 *Treatment*

Treatments took place over four consecutive weeks and began on the second week of the study (see Table 3-1). Each of the four treatment sessions lasted approximately 20 to 30 minutes in length and consisted of a single question elicitation task. The four treatment tasks were similar in design to the pre-test task. For each treatment task, classes were divided into two teams and students in each team were asked to produce questions using vocabulary cards they selected from a game board. The game boards were divided into two sections (one per team), so that each team had its own grid of vocabulary cards. Each grid contained all of the vocabulary needed for a team to complete their task; Team A would only choose cards from Grid A and Team B would only choose cards from Grid B. Once a student asked a question, the vocabulary card was taken back to the student's team where the students worked together to finish the task.

The first and third treatment tasks (Weeks 2 and 4, respectively) consisted of a crossword puzzle; students had to earn all the vocabulary cards for their team and then use the words or phrases on the card to complete the puzzle. Each team received a half-finished crossword puzzle; the blanks on one team's puzzle corresponded with the given

answers of the opposing team's puzzle. An example of the crossword puzzle is available in Appendix B. The second and fourth treatment tasks (Weeks 3 and 5, respectively) were identical to the other treatment tasks except that students used the vocabulary cards to fill in the blanks of a Story Completion task. Each team was given a written copy of a short story with words or phrases missing. The story incorporated all the words and phrases on the game board (that is, all the words and phrases for both teams). As with the crossword puzzle, the missing words of one team's story were provided for the opposing team (and vice versa). An example of the story completion task is provided in Appendix C. A team won by completing the crossword or story before the opposing team finished their version of the same task.

The treatment tasks differed from the pre-test task in that the teachers provided a single form of feedback to the treatment groups in response to ungrammatical questions (see pages 35-36 for examples of feedback types). Both the nature of the tasks and the teachers' responses to ungrammatical questions were highly repetitive; thus, the feedback types under investigation were provided with a high degree of consistency and were quite intensive. As well, the nature of the treatment tasks was almost identical to the pre- and post-test tasks; consequently, the stage of the questions elicited from the pre- and post-test tasks is an appropriate measure of the feedback provided during the treatments.

In order to correspond with the teachers' regular course syllabi, the vocabulary was different for each of the four treatment sessions. In this way, the vocabulary cards gave students a prompt from a familiar theme from which to produce a question. In addition, the students were free to produce any question at any stage. As there was no penalty for producing an incorrect or ungrammatical question, it is unlikely that students

felt overly anxious about trying to produce higher stage questions. In other words, students were not denied a vocabulary card for producing an unacceptable question, and so were free to produce higher stage questions at the risk of being ungrammatical.

3.2.4 Post-tests

The immediate post-test was administered the day after the last treatment (Week 5). The Tic Tac Toe task used for the pre-test was used with the same vocabulary for the immediate post-test (see Table 3-1). The delayed post-test was administered two weeks after the last treatment (Week 7), and used the same vocabulary and crossword puzzle as the first treatment task. The second post-test was administered in order to check for sustained or delayed feedback effects. The post-test tasks consisted of recycled vocabulary so that the students' questions could easily be compared with their questions at the beginning of the study. As with the pre-test task, teachers did not provide any form of corrective feedback during the post-test sessions.

All of the pre-test, treatment and post-test tasks were completed at the beginning or end of the regular class sessions so as not to interrupt the typical progression of the day's lesson. All of the tasks were designed as games in order to reduce any anxiety students might feel at speaking in front of the whole class. Each task was designed to keep students motivated to complete the task despite the lack of explicit feedback from their teachers about the grammaticality of their utterances. The competitive aspect of the tasks gave students a sense of purpose to the game without the teacher having to provide any feedback outside of the treatment conditions. Indeed, many of the students became quite competitive and there was often a lot of laughter and joking throughout each of the

tasks. As well, such tasks are quite common in communicative L2 classrooms and are thus characteristic of tasks the subjects may have engaged in during regular class time outside of the research sessions.

3.3 Coding

This section describes how the data were organised before analysis. Each learner was assigned an identification number and each question was coded according to two important characteristics of question development: developmental stage and grammaticality. The first section describes the process of categorising questions into the appropriate stages of development suggested by Pienemann, Johnston, and Brindley (1988). The second section outlines the coding procedures for measuring a student's possible improvement between the pre-test and the two post-tests. The third section describes the categorisation of questions according to grammaticality. The fourth section describes the reliability of the coding. Approximately 25% of the questions were reviewed by a second coder to ensure inter-coder reliability.

3.3.1 Developmental Stages

Questions were categorised into one of the stages of question development. Table 3-3 provides examples for each developmental stage.

Table 3-3

Developmental Stages and Question Structures

Stage	Structure	Example
1	Single Words	“Cat?” “Why?”
2	Canonical Word Order	“You have a cat?” (Rising Intonation)
3	Fronting	“Where the cats are?” “Do you have a cat?”
	Wh / Do / Q-word	“Can I ride bicycle?” “Is you are happy?”
4	Y/N with aux inversion	“Have you drawn the cat?”
	Wh- with copula BE (Pseudo-Inversion)	“Where are the cats?” “Is the child climbing the tree?” “Have you ever seen a cat stuck in a tree?”
5	Q-word→Aux/modal →subj	“What does your cat eat?” “Why have you left?”
6	Cancel Inv, Neg Qs, Tag Qs	“Can you see what time it is?” “Isn’t your cat pretty?”

**This table is based on Pienemann and Johnston (1987) and is adapted from Spada and Lightbown (1993) and Mackey (1999).*

Questions were coded according to the occurrence and the position of structural elements that characterise each stage. Thus, a question beginning with a *Wh*-word followed by the copula *BE* (i.e. *What is a squirrel?*) was coded as a Stage 4 question even if it was otherwise ungrammatical (i.e. *What is squirrel?*). In contrast, a question was only considered to be characteristic of a particular stage if the components were appropriate for that particular question. For example, Stage 5 questions require a *Wh*-word followed by an auxiliary and a subject (*What do raccoons eat?*). A question like *what does a seal*

colour? was coded as a Stage 4 question (*Wh*-word followed by the copula *BE*) and *not* as a Stage 5 question because the auxiliary *do* is not the correct choice of auxiliary.⁷ In addition, a student's overall stage was determined as the average of the sum of stages the student had produced in a single task. For example, if a student produced one Stage 3 question and one Stage 4 question, the overall level of questions was coded as Stage 3.

3.3.2 *Improvement*

Each question was also coded according to the type of test during which each utterance was produced (pre-test, post-test 1, and post-test 2) and the treatment group to which the speaker was assigned (control, recast, elicitation and recast + prompt). The students' overall performance on each of the post-tests was compared with their performance on the pre-test. Numerical values were assigned to indicate the difference in levels across tests, for example, if a student maintained a Stage 3 level on a post-test, a value of 0 was given under the heading *Improve*. If a student's overall performance had *increased*, a value of 1 was designated.

3.3.3 *Grammaticality*

Coding a question's overall grammatical accuracy separately from the structural level of development provides another index of a student's level of question development in relative isolation from other problems such as subject-verb agreement and appropriate use of determiners. Although the stages of question development are the main focus of this study, the factor of grammaticality might provide a glimpse into a student's question

⁷ *What is a seal's colour?* or *What colour is a seal?* are the most likely target forms that the student was trying to produce.

development within a single stage. First, questions were categorised as being either grammatical (*Yes*) or ungrammatical (*No*). Any type of error, whether an incorrect use of a determiner or a serious syntactic problem, rendered a question to be coded as ungrammatical. In other words, a student's production of Stage 3 questions across all measures (both pre- and post-tests) does not necessarily mean that the student has not developed; it might be possible to observe question development in terms of increased grammaticality over time. Table 3-4 provides an example of the coding:

Table 3-4

Sample of Coding: Grammaticality and Stage Level⁸

Test	ID	Group	Question	Grammatical	Stage	Overall Stage	Post-test Improvement
Pre-test	S36	Elicitation	Confusion. Are you confusion?	No	4	3	N/A
Pre-test	S36	Elicitation	Ad. Do you know- do you know good ad?	No	3		N/A
Post-test 1	S36	Elicitation	Ah, can you- can you owe to me?	No	4	3	0
Post-test 1	S36	Elicitation	Seeing someone. ++ When you see someone? ⁹	No	3		
Post-test 2	S36	Elicitation	Where-ah, how-how did you get harm?	No	5	4	1
Post-test 2	S36	Elicitation	Which- which ad is-+ can I say touched you?	No	4		

⁸ This table has been modified to fit the document format.

⁹ ++ indicates a pause of several seconds. A single + indicates a pause of only one or two seconds.

3.3.4 *Reliability*

I transcribed all of the recordings made of the pre- and post-test sessions and categorised each question as belonging to one of five stages of question development. In order to determine the reliability of the coding, 25% of the data was randomly selected and coded by another applied linguistics graduate student. The second coder categorised each question for grammaticality and developmental stage. The rate of agreement for stage level was 91%. In other words, of the 125 questions judged by the second coder, 114 of the judgements were identical to my own. For grammaticality, the rate of agreement between the two coders was 97% (the coders agreed on 121 of the 125 questions). Any differences in coding were resolved through negotiation. That is, we discussed our reasoning behind our initial categorisations before coming to a consensus for each question. As a consensus was made in all cases, all of the questions collected from the 53 participants were analysed.

3.4 **Statistical Analyses**

In this study, interactional feedback is the independent variable and the level of question development is the dependent variable. The three treatment groups and the control group represent the four layers of the independent variable. The first step of the analyses of this data involved descriptive statistics. For each of the four conditions, the frequencies of Grammaticality (Yes/No) and Stage Level (1-5) were counted for each of the tests (pre- and post-tests). In addition, descriptive statistics were also run to find and compare the frequencies and means of improvement among the groups who did not

receive any feedback and the control group as well as between the treatment groups.

Second, one-way analyses of variance (ANOVA) and analyses of covariance (ANCOVA) were used to compare the means both *between* each of the four groups as well as to compare means *within* groups across the three tests.

Chapter Four

ANALYSES AND RESULTS

The purpose of this section is to display the results of the analyses that have been used for this research. The data and results presented in this section represent the data observed from the 53 students who completed each of the pre- and post-test tasks. The sections in this chapter are organised according to the four research questions that motivate this study. Section 4.1 provides the descriptive statistics and analyses regarding stage level, improvement levels, and grammaticality for the control group and the treatment groups as a whole. Section 4.2 presents the relevant data to compare the recast, elicitation and recast + prompt groups in order to investigate whether learners who receive recasts progress to a higher level of question development than learners who receive elicitation. Section 4.2 also provides the necessary data to examine the third research question, whether recast in combination with elicitation is more effective than only recast or only elicitation. The analyses and results presented in Section 4.2 also compare students' performance on each of the pre- and post-test tasks to determine whether effects of either form of interactive feedback are delayed or maintained over time. Finally, Section 4.3 provides a summary of the main findings presented in this chapter.

4.1 Treatment and Control Groups

The first research question investigates whether learners who receive interactional feedback progress to higher stages of question development than students who do not receive interactional feedback. The analyses and results pertaining to this question have been subdivided into three sections. Section 4.1.1 presents the results and analyses of the overall stages produced by the control group and the treatment groups. Section 4.1.2 presents the results and analyses of the students' improvement levels, and Section 4.1.3 presents similar information about the grammaticality of the students' questions.

4.1.1 Overall Stages by Control and Treatment Groups

The first step of the data analysis compared the overall performance of the control group ($n=15$) to the overall performance of the three treatment groups ($n=38$) as a whole. Table 1 displays the overall stages of questions produced by the control group and the treatment groups on the pre-test and immediate post-test tasks. The numbers presented above the percentages in each row represent the number of students who produced questions at each overall stage, and the final column displays the total number of students in the control and treatment groups on each test.

Table 4-1

		<i>Overall Stages of Control and Treatment Groups</i>				
		Overall Stage				
Test Type		2	3	4	5	Total
Pre-test	Control	0	1	7	7	15
		.0%	6.7%	46.7%	46.7%	100%
	Treatment	1	19	13	5	38
		2.6%	50.0%	34.2%	13.2%	100%
	Total	1	20	20	12	53
Post-test 1	Control	0	7	6	2	15
		.0%	46.7%	40.0%	13.3%	100%
	Treatment	0	19	18	1	38
		.0%	50%	47.4%	2.6%	100%
	Total	0	26	3	3	53
		.0%	49.1%	5.7%	5.7%	100%

Table 4-1 shows that the control group produced a higher percentage (46.7%) of Stage 5 questions on the pre-test than the three treatment groups combined (13.2%). In addition, the lowest stage produced by the control group was Stage 3 (6.7%), while the lowest score for the treatment groups was Stage 2 (2.6%). An alpha level of .05 was used for all statistical tests, and a Chi-square test found this difference in performance to be significant ($p = .005$)¹⁰. This indicates that the students in the control group were performing significantly better than the students in the treatment groups at the beginning of the study; the control group produced more developmentally advanced questions than the treatment groups.

¹⁰ The exact probability value of the likelihood ratio was used for all Chi-square tests.

Both the control group and the treatment groups produced a lower percentage of Stage 5 questions on post-test 1 than on the pre-test (control group = 13.3%, treatment group = 2.6%). However, the treatment groups produced a higher percentage (47.4%) of Stage 4 questions than the control group (40%). There is no significant difference in the stages of questions between the control group and the treatment groups on post-test 1 ($p = .358$). This finding suggests that the students produced similar percentages of Stage 3, Stage 4 and Stage 5 questions; neither group of students produced more developmentally higher stage questions than the other on post-test 1.

4.1.2 *Improvement by Control and Treatment Groups*

The previous section provided descriptive statistics to show how many questions students produced at each developmental stage. However, it is difficult to determine how many students in the control and treatment groups were actually at a higher developmental stage on post-test 1. For example, it is impossible to determine whether the increased percentage of Stage 4 questions on post-test 1 by the treatment groups is due to developmental improvement from Stage 3 questions, or from a decline in Stage 5 questions. This section analyses whether any of the students had *improved* on post-test 1. For this purpose, *Improvement* is operationalised as an increase in a student's overall stage. For example, the overall question stage of each student on post-test 1 was compared with their overall stage on the pre-test. Thus, if a student maintained a Stage 3 level on post-test 1, then no improvement in overall stage was attained. If a student's overall performance had *increased* from Stage 3 to Stage 4, then the student was considered to have improved (see Section 3.3.2 for a detailed description of coding). The

total number of students who improved in this manner was counted. Figure 1 displays the total percentage of students who improved in the control and treatment groups on post-test 1.

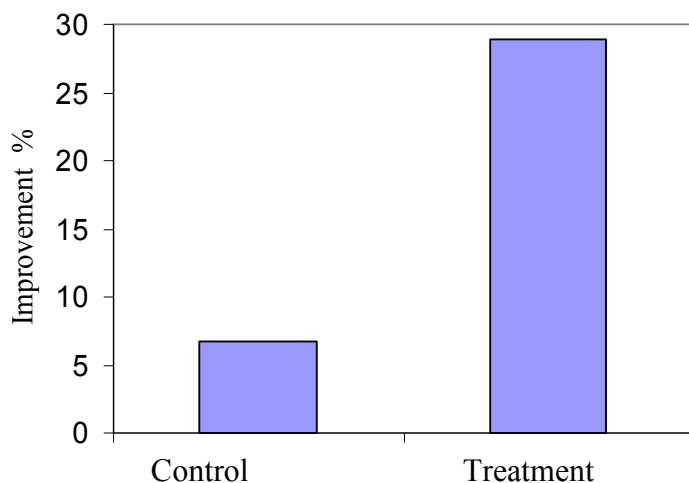


Figure 1 *Overall Improvement of Control and Treatment on Post-test 1*

Figure 1 illustrates that more students in the treatment groups (28.9%) showed improvement on post-test 1 than in the control group (6.7%). Table 4-2 presents this data along with the total number and percentage of students who showed improvement and the total number and percentage of students who did not improve for the control and treatment groups on post-test 1.

Table 4-2

Improvement of Control and Treatment on Post-test 1

		Post-test 1		
		Not Improved	Improved	Total
Control vs. Treatment	Control	14	1	15
		93.3%	6.7%	100%
	Treatment	27	11	38
		71.1%	28.9%	100%
Total		41	12	53
		77.4%	22.6%	100%

Table 4-2 shows that 11 of the 38 students in the treatment groups had improved on post-test 1, while this was true of only 1 of the 15 students in the control group. However, a Chi-square test showed that this difference was only marginally significant ($p = .057$), which suggests that students who received feedback improved noticeably more than students in the control group. These findings indicate that even though students in the control group produced more Stage 5 questions than students in the treatment groups, the students who received feedback improved their question production more than the students who did not receive feedback.

4.1.3 Grammaticality of Control and Treatment Groups

It is possible that students were developing their knowledge of English question structures without having produced higher-stage questions. In order to examine students' intra-stage development, the total number and percentage of grammatical and ungrammatical questions produced by the control group and the treatment groups was

calculated. Table 4-3 displays the number and percentages of grammatical and ungrammatical questions for the pre-test and post-test 1.

Table 4-3

Grammaticality of Control and Treatment Groups

Test Type		Grammatical		
		No	Yes	Total
Pre-test	Control	34	2	36
		94.4%	5.6%	100%
	Treatment	59	34	93
		63.4%	36.6%	100%
	Total	93	36	129
Post-test 1		72.1%	27.9%	100%
	Control	30	3	33
		90.9%	9.1%	100%
	Treatment	48	42	90
		53.3%	46.7%	100%
	Total	78	45	123
		63.4%	36.6%	100%

Table 4-3 shows that only 5.6% of the questions produced by students in the control group were grammatical, while 36.6% of the questions delivered by the treatment groups on the pre-test were grammatical. A Chi-Square test showed there was a significant difference ($p = .00$) between the control and treatment groups on the pre-test, suggesting that students in the treatment groups were more grammatically accurate than those in the control group at the beginning of the study. There was also a significant difference ($p = .001$) between the control group and the treatment groups on post-test 1, with students in the treatment groups producing the higher percentage of grammatical

questions. The data displayed here shows that 46.7% of the treatment students produced grammatical questions, while only 9.1% of the control students' questions were grammatical. Thus, 10.1% of the students who received feedback improved in terms of grammaticality, while 3.5% of the students who did not receive feedback improved the grammaticality of their questions (as calculated by subtracting the percentages on the pre-test from those on the post-test).

This section has provided the overall stage production, improvement and grammaticality of students in the control and combined treatment groups on the pre-test and post-test 1. However, for the remaining research questions to be investigated, it is necessary to examine the performance of each of the three treatment groups. The results and analyses concerning the recast, elicitation and recast + prompt groups are presented in the following section.

4.2 Recast, Elicitation and Recast + Prompt

The previous section presented the analyses and results pertaining to the first research question. In order to investigate the effects of interactional feedback in general, the overall stage, improvement and grammaticality of the questions produced by students in the control group were compared with those produced by the combined treatment groups. This second section provides the results and analyses for the second and third research questions. The second research question investigates whether learners who receive recasts progress to a higher level of question development than learners who receive elicitation. The third research question examines whether the recast + prompt group would perform better than the recast or elicitation groups. As with Section 4.1

above, the following results and analyses are divided into three subsections. Section 4.2.1 presents the overall stages of questions produced by the recast, elicitation and recast + prompt groups on the pre-test and immediate and delayed post-tests. Section 4.2.2 displays the levels of improvement achieved by each group, and Section 4.2.3 shows the grammaticality of questions produced by students in each of these groups.

4.2.1 Overall Stages of Feedback Groups

The first step of the data analysis compared the results of the recast, elicitation and recast + prompt groups. While section 4.1 examined the differences between the control group and the treatment groups as a whole, this section will concentrate on any differences between each specific type of feedback. The data for this analysis is divided into two subsections. Section 4.2.1.1 provides the results and analyses for the pre-test and post-test 1, and Section 4.2.1.2 examines the results for post-test 2 and compares it to those of post-test 1.

4.2.1.1 Overall Stages on Pre-test and Post-test 1

This section investigates the overall stages of questions produced by each of the treatment groups on the pre-test and immediate post-test. Table 4-4 displays the overall stages of questions produced by the recast, elicitation and recast + prompt groups on the pre-test and immediate post-test tasks (see Table 4-1 for the overall stages of questions produced by the control group).

Table 4-4

			Overall Stage				Total
			2	3	4	5	
Pre-test	Group	Recast	0	6	7	3	16
			.0%	37.5%	43.8%	18.8%	100%
		Elicitation	0	4	4	1	9
			.0%	44.4%	44.4%	11.1%	100%
Total	Total	Recast + Prompt	1	9	2	1	38
			7.7%	69.2%	15.4%	7.7%	100%
			1	19	13	12	53
			2.6%	50.0%	34.2%	13.2%	100%
Post-test 1	Group	Recast	0	5	10	1	16
			.0%	31.3%	62.5%	6.3%	100%
		Elicitation	0	4	5	0	9
			.0%	44.4%	55.6%	.0%	100%
Total	Total	Recast + Prompt	0	10	3	0	13
			.0%	76.9%	23.1%	.0%	100%
			0	19	18	1	38
			.0%	50.0%	47.4%	2.6%	100%

All three of the feedback groups produced a higher percentage of Stage 5 questions on the pre-test than on post-test 1. On the pre-test, 18.8% of the questions asked by the recast group were at Stage 5, while 11.1% were Stage 5 for the elicitation group, and 7.7% were Stage 5 for the recast + prompt group. As well, the only group to produce a Stage 2 question (the lowest level produced by any student across all three test tasks) was the recast + prompt group (7.7%). The lowest stage of question produced by the recast and elicitation groups was Stage 3. On post-test 1, the lowest level of question produced by any group was Stage 3. Only the recast group produced Stage 5 questions

on post-test 1, with the recast group producing 6.3% of their questions at this advanced level. Of the total questions produced by the recast group, 62.5% were Stage 4 questions, while 55.6% of the questions produced by the elicitation group were at this stage. The recast + prompt group produced the least percentage of Stage 4 questions (23%). In addition, while the recast group produced fewer Stage 3 questions¹¹ on post-test 1, the elicitation group produced exactly the same percentage of Stage 3 questions (44.4%), and the recast + prompt group produced more Stage 3 questions on post-test 1 (77%) than on the pre-test (69%).

In addition to percentages, means for overall stage levels were also calculated in order to determine whether there were any significant differences in overall stage among the three treatment groups on the pre-test and post-test 1. Table 4-5 displays the number of questions produced by each group, the mean stage of questions as well as the highest and lowest stages of questions produced by each group on the pre-test and post-test 1.

¹¹ Recall that Stage 3 was the lowest level produced by the recast, Elicitation and control groups.

Table 4-5

Number and Mean of Overall Stages on the Pre-test

Group	Test	N of Questions	Mean Stage	Lowest Stage	Highest Stage
Control	Pre-test	36	4.40	3	5
	Post-test 1	33	3.66	3	5
Recast	Pre-test	37	3.81	3	5
	Post-test 1	35	3.75	3	5
Elicitation	Pre-test	29	3.66	3	5
	Post-test 1	25	3.55	3	4
Recast +	Pre-test	27	3.23	2	5
Prompt	Post-test 1	30	3.24	3	4

On the pre-test, the control group showed the highest mean for Stage level (*mean* = 4.40), with the recast (*mean* = 3.81) and elicitation (*mean* = 3.66) groups producing the second and third highest means, respectively. However, on post-test 1, the control, recast and elicitation groups all had lower means (*means* = 3.66, 3.75 and 3.55, respectively) than they did on the pre-test. The recast + prompt group showed a slightly higher mean on post-test 1 (*mean* = 3.24) than on the pre-test (*mean* = 3.23). This indicates that the recast + prompt group was the only group to advance to a higher overall stage of questions between the pre-test and post-test 1.

Having calculated the means, a one-way Analysis of Variance (ANOVA) was used to compare the means for overall stage levels produced by each group. An ANOVA was used for this because there were more than two groups to be compared. The ANOVA was used on the pre-test data first to examine which of the feedback groups

were significantly different from the control group (recall that students in the control group produced significantly more advanced stage questions than students in the treatment groups on the pre-test (See Section 4.1.1)). The One-way ANOVA found a significant difference between groups on the pre-test, $F(3, 49) = 6.553$, $p = 0.001$. Next, a post-hoc Tukey test was used in order to determine *which* of the treatment groups were different. There was a significant difference found between the control group and the recast + prompt group ($p = 0.00$), with the control group producing more developmentally advanced questions. Table 4-6 shows the mean differences between groups.

Table 4-6

Mean Difference Between Groups on the Pre-test

(I) Group	(J) Group	Mean Difference (I-J)	Sig.
Control	Recast	.59	.108
	Elicitation	.73	.078
	Recast + Prompt	1.17	.000**
Recast	Elicitation	.15	.959
	Recast + Prompt	.58	.135
Elicitation	Recast + Prompt	.44	.489

** *Alpha level for significance is less than .01*

There was no significant difference between any of the other treatment groups and the control group on the pre-test. Similarly, there was no significant difference between any of the treatment groups on the pre-test. This indicates that the difference found between the combined treatment groups and the control group was due only to the difference between the control group and the recast + prompt group. Otherwise, these

findings suggest that participants in the four conditions were at the same level of question development. Because the questions produced by the control group were significantly more advanced than those produced by the recast + prompt group, it was necessary to control for the influence this extra variable might have on the results for post-test 1. Thus, an Analysis of Covariance (ANCOVA) was used on the post-test 1 data to partial out the difference between the control and recast + prompt groups on the pre-test. The overall stage of questions produced by each group on post-test 1 was entered as the dependent variable, with the same data from the pre-test entered as a covariate. In other words, the initial difference on the pre-test was taken into account so that any differences found between groups on post-test 1 would be statistically robust. There was no significant difference between the control group and any of the treatment groups on post-test 1 $F(3, 48) = 2.108, p = 0.112$. See Table 4-5 for descriptive statistics of the pre-test and post-test 1.

Recall that there was no significant difference between the three feedback groups on the pre-test. Therefore, the post-test 1 results of the three groups who received feedback were compared with a second One-way ANOVA. As it had already been determined that there was no significant difference between the control group and the treatment groups (see previous paragraph), the control group was excluded from this test. A significant difference was found between the treatment groups $F(2,35) = 3.569, p = .0039$. Again, a post-hoc Tukey test was used to determine which groups were significantly different. The recast group was found to be significantly better than the recast + prompt group on post-test 1 ($p = 0.03$). There was no significant difference between the recast group and the elicitation group or between the elicitation and recast +

prompt groups on post-test 1. Table 4-7 displays the mean differences between the three feedback groups on post-test 1.

Table 4-7

Mean Difference Between Feedback Groups on Post-test 1

(I) Group	(J) Group	Mean Difference (I-J)	Sig.
Recast	Elicitation	.19	.647
	Recast + Prompt	.52	.030*
Elicitation	Recast + Prompt	.32	.334

* *Alpha level of significance is less than .05.*

These results indicate that all of the groups produced questions at similar stages of development on post-test 1 except for the recast group, who produced significantly more high-stage questions than the recast + prompt group (see Table 4-4 for the overall stages of each group).

4.2.1.2 Overall Stages on Post-test 2

This section considers whether the effects of either form of interactional feedback are maintained over time. It displays the analyses and results of the overall stages produced by students in each of the four conditions on post-test 2. Participants' production on this delayed post-test task is compared to their performance on the immediate post-test (post-test 1) and the pre-test. Table 4-8 displays the overall stages of questions produced by each group on post-test 2.

Table 4-8

Overall Stages on Post-test 2

		Overall Stage			
		3	4	5	Total
Group	Control	3	12	0	15
		20.0%	80.0%	.0%	100%
	Recast	5	12	0	17
		29.0%	71.0%	.0%	100%
	Elicitation	2	6	1	9
		22.0%	67.0%	11.0%	100%
	Recast + Prompt	7	6	0	13
		54.0%	46.0%	.0%	100%
Total		17	36	1	54
		31.0%	67.0%	1.9%	100%

Table 4-8 shows that students in the elicitation group produced the most developmentally advanced questions on post-test 2, with 11% of their questions at the Stage 5 level. No other students produced Stage 5 questions on this delayed task. The control group produced the highest percentage of Stage 4 questions (80%). Of the questions produced by students in the recast group, the majority were Stage 4 questions (71%). This is also true of the elicitation group, where 67% of the questions were Stage 4. The recast + prompt group produced the least percentage of Stage 4 questions (46%), with 54% of their questions at the Stage 3 level. The lowest Stage produced by any group was Stage 3. These results show that the students' question production on post-test 2 was not consistent with their production on post-test 1: the control group produced fewer Stage 5 questions on each succeeding post-test than on the pre-test, the elicitation group increased production of Stage 5 questions on post-test 2, the recast group increased

production of Stage 4 questions on post-test 2 but did not produce any Stage 5 questions and while the students in the recast + prompt group did not produce any Stage 5 questions on either post-test, their production of Stage 4 question increased with each succeeding test.

The next step of the analysis was to determine whether any of the differences observed between groups were statistically significant. As with post-test 1 above, the mean stages of questions produced by each group on post-test 2 were compared. Table 4-9 presents the mean overall stage for each group on post-test 2 as well as the highest and lowest stage of questions.

Table 4-9

Number and Mean of Overall Stages on Post-test 2

Group	N of Questions	Mean Stage	Lowest Stage	Highest Stage
Control	36	3.80	3	4
Recast	38	3.69	3	4
Elicitation	31	3.89	3	5
Recast + Prompt	35	3.46	3	4

Because the control group produced significantly more high stage questions on the pre-test than the recast + prompt group (see Section 4.2.1.1), an ANCOVA was used to filter out this initial difference. No significant difference was found between the four groups $F(3, 48) = 1.141$, $p = .067$. This suggests that the control group was producing questions at a similar level to the treatment groups. For the final step of this part of the analysis, the control group was once again excluded from a One-way ANOVA that was used to compare the overall stages of the three treatment groups. There was no

significant difference found between groups $F(2, 35) = .719$, $p = .494$. In other words, the students in the feedback groups were producing the same overall stage of questions on post-test 2.

4.2.2 Improvement of Treatment Groups

In the previous section, the overall stages of questions produced by each group were compared for the pre-test and post-test 1. The next step in the analysis examined how many of the students in each of the four conditions improved. As with Section 4.1.2, it is difficult to see from the analyses of overall stages alone which students have improved between the pre-test and post-test 1. For example, the observed increase in Stage 4 questions on post-test 1 by the recast group could be due to students advancing from Stage 3, or from a regression from Stage 5. Thus, this section investigates whether any of the students had improved on each of the post-test tasks. This section is divided into two subsections. Section 4.2.2.1 reveals whether any of the students had improved on post-test 1, while Section 4.2.2.2 examines students' improvement on post-test 2.

4.2.2.1 Improvement of Treatment Groups on Post-test 1

Each student's overall stage on post-test 1 was compared with their overall stage on the pre-test (see Section 4.1.2 for a more detailed explanation of how improvement is operationalised in this study). Figure 2 displays the percentage of students who improved

for each group on post-test 1 (the data for the control group is repeated here for convenience).

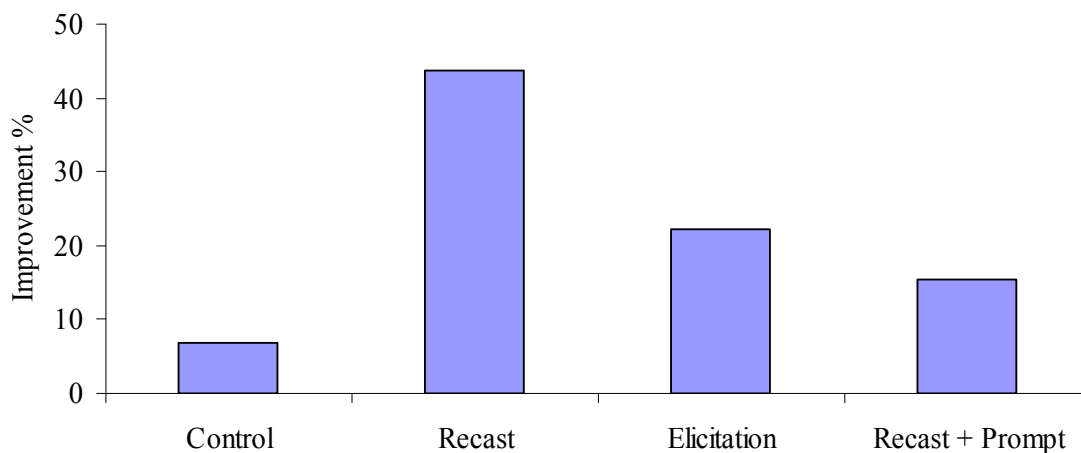


Figure 2 *Overall Improvement of All Groups on Post-test 1*

All of the groups who received feedback showed more improvement in overall stage production than the control group, with the recast group showing the highest level of improvement (43.8%). Table 4-10 provides further details of the data presented in Figure 2 (above), including the total number and percentage of students who showed improvement and who did not improve for each of the four conditions on post-test 1.

Table 4-10

Improvement of All Groups on Post-test 1

		Post-test 1		
		Not Improved	Improved	Total
Group	Control	14	1	15
		93.3%	6.7%	100%
	Recast	9	7	16
		56.3%	43.8%	100%
	Elicitation	7	2	9
		77.8%	22.2%	100%
	Recast + Prompt	11	2	13
		84.6%	15.4%	100%
Total		41	12	53
		77.4%	22.6%	100%

More than twice as many students in the recast group showed improvement (43.8%) than students in the elicitation group, who had the second highest percentage of improvement (22.2%). While students in the recast + prompt group did not improve as much as the other feedback groups, they still improved more than the control group (recast + prompt = 15.4%, control = 6.7%). While a Chi-square test found no statistically significant difference between groups ($p = .081$), the results show a positive trend for feedback effects.

4.2.2.2 *Improvement of Treatment Groups on Post-test 2*

This section presents the analyses and results of the students' improvement levels on post-test 2, and compares them with the improvement levels of post-test 1. This

section will focus on the level of improvement achieved by each group between the pre-test and post-test 2. Figure 3 displays the percentage of students in each group who improved on post-test 2.

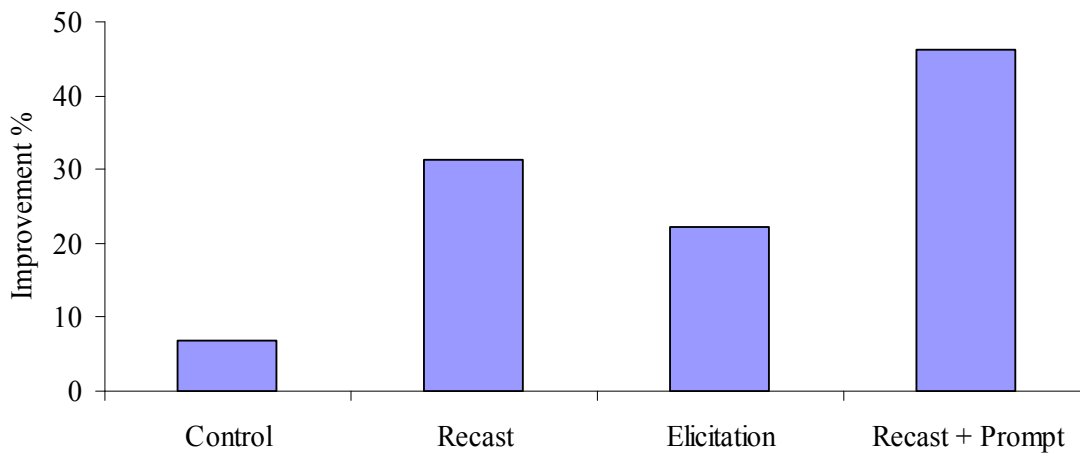


Figure 3 *Overall Improvement on Post-test 2*

All of the treatment groups showed more improvement than the control group on post-test 2, with the recast + prompt group showing the highest level of improvement (46.2%). Table 4-11 provides a detailed presentation of the frequency data.

Table 4-11

Improvement on Post-test 2

		Post-test 2		
		Not Improved	Improved	Total
Group	Control	14	1	15
		93.3%	6.7%	100%
	Recast	11	5	16
		68.8%	31.3%	100%
	Elicitation	7	2	9
		77.8%	22.2%	100%
	Recast + Prompt	7	6	13
		53.8%	45.2%	100%
Total		39	14	53
		73.6%	26.4%	100%

A comparison of the results between the two post-tests (see Section 4.2.2 for the improvement results for post-test 1), reveals that the recast + prompt group also improved more on post-test 2 (46.2%), suggesting delayed positive effects for the recast + prompt group. Students in the recast group showed the second highest improvement on post-test 2 (31.3%); however, this is lower than their improvement on post-test 1 (43.8%). Thus, there seems to be no long-term or delayed effects for the recast group. The elicitation group maintained an improvement level of 22.2% on both post-tests, suggesting some long-term effects. Similarly, the control group maintained a 6.7% improvement level on post-test 2. A Chi-square test was used to determine any significant difference between groups. As with post-test 1, no statistically significant difference was found ($p = .090$). Nevertheless, the finding that the control group showed the least amount of improvement

(6.7%) on both immediate and delayed post-tests indicates a positive effect of feedback.

Figure 4 presents a summary of the improvement values of each group on post-test 1 and post-test 2.

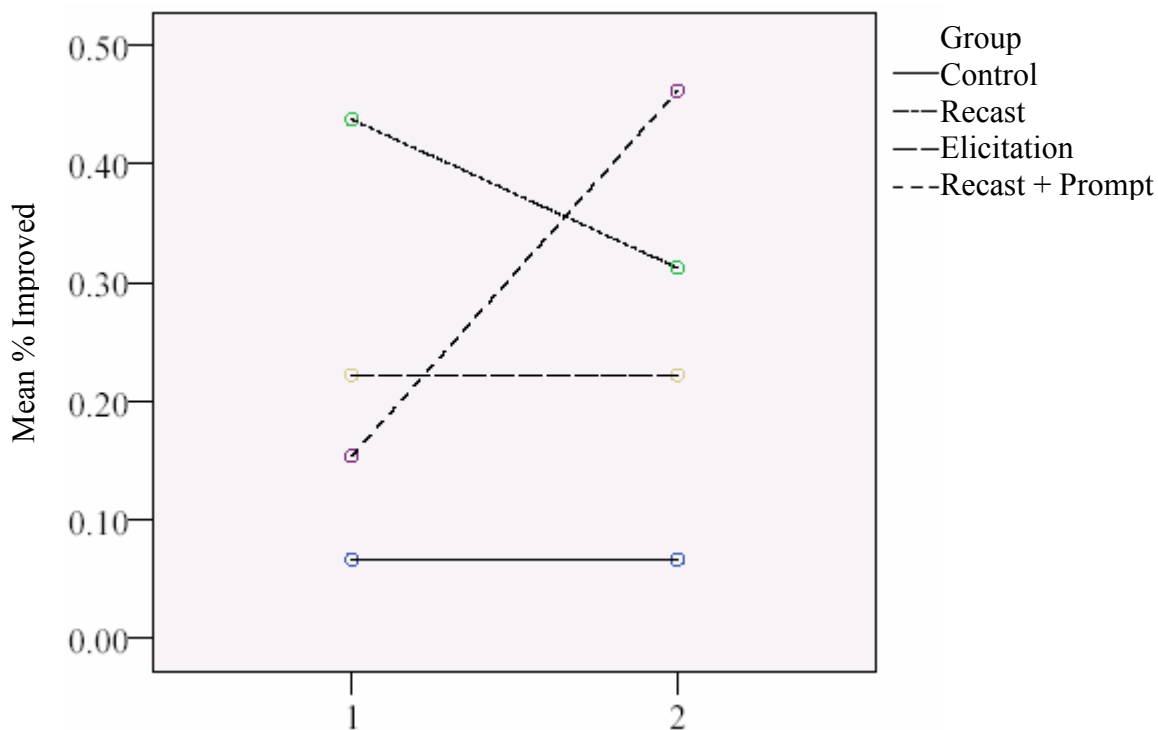


Figure 4 *Immediate and Delayed Improvement*

In sum, the improvement values are highly varied both among groups and across post-tests. The control group showed little improvement on the post-tests. The recast group achieved the greatest improvement levels on post-test 1, but did not maintain their improvement on post-test 2. In contrast, the recast + prompt group showed the least improvement of the three treatment groups on post-test 1, but achieved the highest improvement score of any group on post-test 2. In addition the elicitation group maintained the same level of improvement on post-test 1 and post-test 2.

4.2.3 Grammaticality of Treatment Groups

In Section 4.1, the grammaticality of the questions produced by the control and treatment groups was examined. The results suggest that students in the treatment groups were significantly more grammatically accurate than students in the control group on both the pre-test and post-test 1. This section will compare the grammaticality of questions produced by the recast, elicitation and recast + prompt groups on both the immediate and delayed post-tests. Thus, this section is divided into two subsections. The first section displays and analyses the results of post-test 1, and the second section provides the same information for post-test 2.

4.2.3.1 Grammaticality of Treatment Groups on Pre-test and Post-test 1

This section displays the analyses and results concerning the grammatical accuracy of each group who received feedback on the immediate post-test. Table 4-12 displays the total number and percentage of grammatical and ungrammatical questions produced by students in each of the three treatment conditions on the pre-test and post-test 1 (see Table 4-3 for the grammaticality of the control group).

Table 4-12

			<i>Grammaticality of All Groups on the Pre-test and Post-test 1</i>		
Test Type			Grammatical		
	Group	Recast	No	Yes	Total
Pre-test		Recast	24	13	37
			64.9%	35.1%	100%
		Elicitation	19	10	29
			65.5%	34.5%	100%
Pre-test		Recast + Prompt	16	11	27
			59.3%	27.9%	100%
		Total	93	36	129
			72.1%	27.9%	100%
Post-test 1		Recast	21	14	35
			60.0%	40.0%	100%
		Elicitation	14	11	25
			56.0%	44.0%	100%
Post-test 1		Recast + Prompt	13	17	30
			43.3%	56.7%	100%
		Total	78	45	123
			63.4%	36.6%	100%

Table 12 shows that the three groups who received feedback produced more grammatical questions on post-test 1 than the pre-test. The recast + prompt group produced the highest number of grammatical questions on both the pre-test (40.7%) and post-test 1 (56.7%). Moreover, the recast + prompt group also improved the most in terms of grammaticality (16%) in that students produced more grammatical questions on post-test 1 than they did on the pre-test. The elicitation group showed the second highest improvement in grammaticality (9.5%), with the recast group improving by 4.9%.

In section 4.1.3, a Chi-square test showed that students in the treatment groups as a whole produced significantly more grammatical questions than students in the control group. Thus, the next step of the analysis was to compare the mean percentage of grammatical questions produced by the individual treatment groups and the control group on the pre-test to determine *which* groups were different. Table 4-13 displays the mean percentage of grammatical questions of each group, as well as the lowest and highest percentages produced on the pre-test and post-test 1.

Table 4-13

Mean % of Grammatical Questions

Group	Test	Mean % Grammatical	Lowest %	Highest %
Control	Pre-test	6.67	0.0	50.00
	Post-test 1	10.00	0.0	50.00
Recast	Pre-test	33.31	0.0	100.00
	Post-test 1	40.63	0.0	100.00
Elicitation	Pre-test	28.67	0.0	50.00
	Post-test 1	44.22	0.0	100.00
Recast + Prompt	Pre-test	46.08	0.0	100.00
	Post-test 1	55.77	0.0	100.00

Table 4-13 shows that the recast + prompt group had the highest mean percentage of grammatical questions on the pre-test (46.08%). The recast and elicitation groups both produced a higher mean percentage (33.31% and 28.67%, respectively) than the control group (6.67%). In order to ascertain whether these differences in grammaticality were significant, the mean percentages of grammatical questions produced by each group on

the pre-test were compared with a One-way ANOVA. A significant difference was found between groups ($F(3, 49) = 3.779, p = .016$). A post-hoc Tukey test was used to determine which groups were significantly different from the control group. The recast + prompt group was the only group whose production of grammatical questions was significantly higher than the control group ($p = .011$). Table 4-14 shows the mean differences between groups on the pre-test.

Table 4-14

Mean Difference in Grammaticality on the Pre-test

(I) Group	(J) Group	Mean Difference (I-J)	Sig.
Recast + Prompt	Control	39.41	.011*
	Recast	12.76	.708
	Elicitation	17.41	.593
Recast	Control	26.65	.106
	Elicitation	4.65	.985
Elicitation	Control	22.00	.368

There was no significant difference between any of the other groups. This means that all of the groups who received feedback were at a similar level of accuracy at the time of the pre-test. These results also indicate that the control group was comparable to the recast and elicitation groups.

An ANCOVA was run on the grammatical data for post-test 1 to account for the difference between the control and recast + prompt groups on the pre-test. Similar to the ANCOVA for overall stage production (see Section 4.2.1), the mean percentage of grammatical questions produced by each group was the dependent variable and the same

data for the pre-test was the covariate. A significant difference was found between groups $F(3, 48) = 3.022, p = .039$. The next step was to determine which groups were significantly different on post-test 1; thus, a Pairwise Comparisons was used to compare the means of each group. Table 4-15 shows the mean differences between groups.

Table 4-15

Mean Difference in Grammaticality on Post-test 1

(I) Group	(J) Group	Mean Difference (I-J)	Sig.
Recast	Control	28.24	.040*
Elicitation	Control	32.26	.041*
	Recast	4.01	.787
Recast + Prompt	Control	42.24	.006**
	Recast	14.00	.300
	Elicitation	9.99	.525

Table 4-15 shows that the grammaticality of all the treatment groups was significantly different from that of the control group. Based on the descriptive data displayed in Table 4-12 above, it is clear that the three treatment groups produced more grammatical questions than the control group. The recast + prompt group was significantly more grammatically accurate than the control group ($p = .006$) on post-test 1. As well, while the recast, elicitation and control groups were at a similar level of grammatical accuracy on the pre-test, the recast ($p = .040$) and elicitation ($p = .041$) groups were also more grammatically accurate than the control group. However, there was no significant difference between any of the three feedback groups on post-test 1.

4.2.3.2 Grammaticality of Treatment Groups on Post-test 2

This section presents the grammaticality of students' questions on post-test 2.

Table 4-16 provides descriptive statistics concerning grammatical and ungrammatical questions produced by each group on post-test 2.

Table 4-16

Grammaticality on Post-test 2

		Grammatical		
		No	Yes	Total
Group	Control	23	13	36
		63.9%	36.1%	100%
	Recast	19	19	38
		50.0%	50.0%	100%
	Elicitation	15	16	31
		48.4%	51.6%	100%
Total	Recast + Prompt	21	14	35
		60.0%	40.0%	100%
		78	62	140
		55.7%	44.3%	100%

Table 4-16 illustrates that the elicitation group showed the highest level of grammatical accuracy on post-test 2 (51.6%). The recast group showed a slightly lower level of grammaticality (50%) than the elicitation group, while producing more grammatical questions than the recast + prompt group (40%) and the control group (36.1%). As with post-test 1, the mean percentages of grammatical questions produced by each of the four groups on post-test 2 were compared. Recall that the recast + prompt group produced significantly more grammatical questions than the control group on the pre-test (see Section 4.2.3). An ANCOVA was used on the grammatical data for post-

test 2 to account for this difference; thus, the percentage of grammatical questions produced by each group on post-test 2 was the dependent variable and the same data for the pre-test was the covariate. There was no significant difference found between the four groups on post-test 2 $F(3, 48) = .869, p = .464$. In other words, the students were producing comparable percentages of grammatical questions on post-test 2. This suggests that the immediate effects (see Section 4.2.3) observed for the treatment groups on post-test 1 were not sustained for the delayed post-test.

4.3 Summary of Results

Chapter four has presented the data used to investigate the four research questions motivating this study. With respect to the first question concerning the effects of interactional feedback, the results suggest that students who received feedback were more likely to produce questions at a higher developmental level on the immediate post-test than students who did not receive feedback. The results also indicate that students who received feedback produced more grammatical questions on the immediate post-test than students in the control group.

The second section investigated whether one form of feedback was more effective than another. The results indicate that while the recast group produced the highest mean stage of questions, the recast + prompt group was the only group to produce a higher mean on post-test 1 than on the pre-test. However, the only significant difference on post-test 1 was found between the recast and recast + prompt groups, with the recast group producing a higher overall stage of questions. In addition, the recast group also showed the most improvement on post-test 1. The elicitation group also improved more

than the recast + prompt group, although there was no significant difference found between any of the groups. In contrast, the recast + prompt group showed the highest level of grammatical accuracy on the immediate post-test; the only group to produce a significantly higher number of grammatical questions than the control group was the recast + prompt group. There was no significant difference in grammaticality between any of the three treatment groups.

As regards the last research question concerning long-term effects of feedback, the results suggest the elicitation group produced the highest stage of question on the delayed post-test. In spite of this, there was no significant difference in overall stage level between any of the groups. With respect to improvement, the recast + prompt group showed the greatest progress on post-test 2. These results suggest some delayed effects for this feedback type, as students improved more on the delayed post-test than on the immediate post-test. The results also indicate that any positive effects of recasts observed on post-test 1 were not maintained on post-test 2. Conversely, the elicitation group maintained the same improvement values across both post-test tasks. Lastly, there was no significant difference between any of the groups in terms of grammaticality on post-test 2, although the elicitation group showed the highest level of grammatical accuracy.

These results illustrate the complexity of measuring the effects of interactional feedback; while there seems to be a trend in favour of positive feedback effects in general, there is no clear winner in terms of the effect of a single type of feedback on question development. Students' performance in the three treatment conditions was highly varied in the production of questions at different developmental stages, in

improvement values across post-tests and in grammatical accuracy. The goal of chapter five is to discuss the implications of these results.

Chapter Five

DISCUSSION

Chapter Four presented the findings concerning the effects of interactional feedback on ESL question development. The goal of this chapter is to discuss the intricacies of these findings and explore possible explanations for the observed outcomes. Chapter Five is divided into five sections with each section dissecting the results and analyses pertaining to one of the four research questions. The first section discusses the positive finding that students who received feedback typically produced higher stage questions with greater grammatical accuracy on the post-test tasks than students in the control group. The second section compares the results of the recast and elicitation groups on the immediate post-test; while students in the recast group showed the highest level of improvement, students in the elicitation group were more grammatically accurate. These findings are partially supported by previous research in the field. The third section compares the results of the recast + prompt group to that of the recast and elicitation groups. This part of the discussion will focus on the surprising finding that the effect of combining recasts and elicitations when providing feedback was less than the observed effects of a single type of feedback on the immediate post-test. Finally, the fourth section focuses on the results of the delayed post-test; students in each of the feedback groups performed quite differently on the delayed post-test than they had on the immediate post-test.

5.1 Feedback Effects

Before examining the effects of individual types of interactional feedback, the goal of the first research question was to investigate the effects of interactional feedback compared to no feedback at all. The data from the three treatment groups were combined and compared as a whole to that of the control group. The first research hypothesis predicted that learners who received interactional feedback would produce questions at a higher stage than learners who did not receive any feedback. The data confirm this hypothesis and are discussed in this section.

First, the results and analyses show that learners in the treatment groups produced higher overall stages of questions than those in the control group on post-test 1. To review, the overall stages of questions produced by students in the control and treatment groups on the immediate post-test were compared with the overall stages of questions students produced on the pre-test. The number of students who produced questions at a higher overall stage on post-test 1 was added and the results found that students who received feedback improved more (albeit marginal significance, $p = .057$) than students who did not receive feedback. This finding is consistent with the literature investigating the effects of interactional feedback as a means of drawing learners' attention to form (Long & Robinson, 1998; Lyster & Ranta, 1997; Mackey & Oliver, 2002).

What makes these results especially noteworthy is the fact that students in the control group were producing questions at a higher developmental stage on the pre-test than students in the treatment groups. This means that while students in the control group may have been more "ready" (Mackey & Philp, 1998) at the time of the pre-test to develop higher stage questions, students in the treatment groups did better than those in

the control group on post-test 1. While students in all four classes were placed at the same intermediate level, students at the English Language Centre are typically placed in classes according to their scores on the placement test (see Appendix A for an example of the SLEP test). Thus, students with lower-intermediate scores would have been placed together while students with scores at a higher-intermediate level would have been placed together. While I never observed the placement test scores of any students, this is offered as a plausible explanation for the initial difference between the students in the control and treatment groups. If this was the case, then the fact that students in the treatment groups improved significantly more on post-test 1 than students who did not receive feedback provides strong support for providing interactional feedback in the L2 classroom.

Second, the first research hypothesis is also supported by the finding that students who received feedback also produced significantly more grammatical questions on post-test 1 than students who did not receive feedback. Again, despite the control group's production of higher-stage questions on the pre-test, it is interesting that more students in the treatment groups *improved* their grammar than students in the control group. For these reasons, this study provides supporting evidence that interactional feedback has positive effects on ESL question development. Furthermore, while other studies that used question development as a measure of acquisition have found similar results in favour of feedback (Mackey, 1999; Mackey & Philp, 1998; Spada & Lightbown, 1999), this study looked at both inter-stage *and* intra-stage development. By looking at the grammaticality of question development, this research allowed a closer look at how feedback affects such development.

While the analysis of overall stage provided marginally significant evidence of feedback effects, the analysis of grammar offers extra support that the provision of feedback had a positive effect on the participants question development. Specifically, these results provide support for Swain's Output Hypothesis (Swain, 1995; 2000). Swain proposes that producing the L2 allows learners to test their hypotheses and thereby increase their fluency in a way that input alone does not provide (Swain, 2000). In other words, engaging in interaction increases the likelihood that learners will notice the differences between their own output and that of the other speaker. This is illustrated by the performance of the students who received interactional feedback in that they produced more grammatical questions on the immediate post-test than they had on the pre-test. As Swain noted, "if learners were not testing hypotheses, then changes in their output would not be expected following feedback" (p. 131). In the case of students in the control group, no such opportunities for noticing or hypothesis testing were provided as the teacher did not provide any feedback in response to their errors.

Having found support for feedback effects and the first research hypothesis, the following section investigates the results pertaining to the overall stage production and grammaticality of each type of feedback.

5.2 Recasts and Elicitations

The previous section discussed the results found for the effects of interactional feedback as measured by overall stage improvement and grammaticality. This section discusses the results and analyses relevant to the second research question, whether learners who received recasts progressed to a higher level of question development than

learners who received elicitation. The corresponding hypothesis predicted that learners who received recasts would progress to a similar level of question development as learners who received elicitation. The results partially support this hypothesis, although the two measures of development provide somewhat different outcomes.

To review the results, there was no significant difference found between the recast group and the elicitation group on post-test 1. In other words, students in both groups were producing questions at similar stages of development and grammaticality. However, learners who received recasts produced slightly higher stage questions on the immediate post-test than learners who received elicitation. While this difference was not statistically significant, the results suggest positive effects of recasts that are not evident for elicitation in terms of overall stage development. On the other hand, learners in the elicitation group produced slightly more grammatical questions on post-test 1 than learners in the recast group. While learners in both groups produced more grammatical questions than learners in the control group on post-test 1, there was no statistical significance in grammaticality between the recast and elicitation groups. These results reflect the mixed findings of previous studies; however, unlike previous research, results of post-test 1 show positive effects for *both* recasts and elicitations as revealed through two measures of development. A discussion of Lyster and Mori's (2006) study concerning interactional feedback and instructional counterbalance sheds some lights on these complex findings.

Their study investigated the immediate effects of explicit correction, recasts, and prompts on learner uptake and repair through teacher-student interactions in elementary-school level French immersion and Japanese immersion classes. Results show a

prevalence of recasts in both instructional settings. There was also a marked difference in uptake and repair patterns as they relate to the different types of feedback, with the highest occurrence of repair a product of prompts in the French immersion setting, and resulting from recasts in Japanese immersion. These apparently divergent results are reflected in the results of this thesis. However, Lyster and Mori discuss two possible explanations for these findings.

First, recasts and prompts elicit different types of learner responses, so it is fundamentally difficult to compare their effects when, in essence, they function rather differently. A learner's repetition of a recast may indicate that the learner has noticed the gap between their initial utterance and the recast; in contrast, a prompt or elicitation often leads the learner to self-repair without providing the learner with positive evidence of the form in question. However, it is still questionable whether a repetition of a teacher's or native-speaker's recast by a learner is a reliable indication that the learner has truly noticed an error, and also whether pushing a learner to self-repair by means of an elicitation leads to successful repair without the provision of positive evidence.

Second, Lyster and Mori propose their counterbalance hypothesis, which asserts that interactional feedback and instructional activities that act as a counterbalance to a communicatively orientated classroom are likely to be more effective than such feedback and activities that more closely resemble its communicative environment. For example, the students in the Japanese immersion class were more exposed to an emphasis on accurate oral production (through reading aloud and choral repetition) than students in the French immersion classrooms; this likely primed the students for repeating recasts. Thus, prompts filled an instructional gap, or counterbalanced a communicative focus in the

French classroom. In the case of the present study, students in both the recast and elicitation groups performed the same task in nearly identical contexts; while both teachers used a communicative approach, the course syllabus included several grammatical structures and corresponding functions. It is possible that the tasks provided learners with a higher exposure to feedback than would have typically occurred in their communicative classrooms and that this “counterbalance” contributed to finding both recasts and elicitations to have been effective, albeit measured in different ways. This suggestion is supported by the finding that students in both the recast and elicitation groups produced significantly more grammatical questions than students in the control group. In other words, students who received recasts or elicitations may have benefited from a counterbalance to the communicative classroom that was absent for the control group.

The hypothesis that recasts would not promote learners to a higher developmental stage than elicitations is only supported by one of the two measures of development employed in this study. The following section discusses the results pertaining to the third research question investigating the effects of recasts in combination with prompts.

5.3 Recasts, Elicitations, or Recast + Prompt

The previous section discussed the individual effects of recasts and elicitation. This section will compare these effects with those of the recast + prompt group, who received both recasts and an elicitation (often occurring in the form of a prompt) to determine whether it is more effective to provide recasts and elicitation together than to

provide only one of these types of feedback. The third research hypothesis predicted that learners who received both recasts and elicitation would produce questions at a higher stage than learners who received only one form of feedback. As with the second research question, the results are not straightforward and possible explanations for the rather confusing findings are discussed. Before discussing the implications of the results, a review of the findings presented in Chapter Four is provided.

The first step of this part of the analyses was to determine which of the three feedback groups was significantly different from the control group at the beginning of the study. The results found a significant difference between the recast + prompt group and the control group in that students in the recast + prompt group were producing questions that were at a lower stage than students in the control group on the pre-test. There was no significant difference between the recast + prompt group and any of the other groups. Considering this difference in performance at the onset of the study, it is interesting that the only students who produced a higher overall mean stage on post-test 1 than on the pre-test were in the recast + prompt group. However, the only significant difference in overall stage production on the immediate post-test was found between the recast group and the recast + prompt group, with students in the recast group producing higher stage questions. In terms of *improvement*, however, there was no significant difference between any of the four conditions on post-test 1. Finally, the measure of grammaticality shows that learners in the recast + prompt group not only produced the highest number of grammatical questions on both tests, but also improved the grammaticality of their questions more than learners in the other groups. While learners in the control group produced significantly higher stage questions than learners in the recast + prompt group

on the pre-test, the converse is true for grammaticality; learners in the recast + prompt group produced significantly more grammatical questions than those in the control group on the pre-test and post-test 1. There was no significant difference in grammaticality between the three treatment groups on post-test 1.

These results do not support the research hypothesis: learners who received both recasts and prompts did *not* advance to higher stages of question production than learners who received only recasts or only elicitation. Studies have shown that recasts typically occur quite frequently in the L2 classroom (e.g., Lyster & Ranta, 1997; Panova & Lyster, 2002), yet the results of their effectiveness in L2 acquisition have not been conclusive. The results of this study prove to be no exception. However, this study provides some evidence that specific conditions may be necessary for recasts and elicitations to be effective. For example, the recasts provided to students in this research were provided intensively during dyadic task-based interaction that explicitly focused on question production. Although recasts may be ambiguous in strictly meaning-oriented tasks, when it is explicitly explained to the students which forms are the focus of the exercise, as was the case for this research, then intensively provided recasts seem to have an important role in the L2 classroom. This finding is especially interesting when recent research into the frequency of recasts in the L2 classroom is considered. In particular, studies by Loewen and Philp (2006) and Lochtman (2002) shed some light on the characteristics and contexts in which recasts are most effective.

In their study of 12 adult ESL classes, Loewen and Philp (2006) investigated the provision and success of recasts and found that both their occurrence and effectiveness were similar to other types of feedback. Although elicitations were related to more

successful uptake than recasts, two thirds of the recasts led to successful uptake. A particularly interesting aspect of their study is the finding that recasts with declarative intonation and prosodic emphasis were more likely to be followed by successful uptake than recasts without such characteristics. This finding suggests that some recasts may be less ambiguous than others and therefore more likely to be interpreted by learners as a corrective move during a communicative interaction. With regards to the results presented in this thesis, Loewen and Philp's study supports the implication that some conditions may be necessary for recasts to be effective. In particular, the explicitness and saliency of the recasts are likely predictors of their effectiveness.

The indication that the explicitness of recasts plays an important role in their potential for successful uptake is also explored in Lochtmann's (2002) study of oral corrective feedback in an analytic L2 classroom. For the purposes of her study, a classroom or task is considered to be analytic when learners are engaged in analysing the L2 in some manner. Lochtmann's experiment took place in Flanders, Belgium, where 15-16 year-old native Dutch speakers were tape-recorded during 12 lessons of German as a foreign language. During this time, the three teachers did not change their teaching methods while their correction methods were being observed. Lochtmann argues that recasts in analytic FLT greatly resemble explicit corrections and are thus quite salient in nature. Recasts were compared with clarification requests, metalinguistic feedback, elicitations and repetitions. Results found that more than half of the total teacher correction moves provided the learners with the opportunity to correct their utterances, indicating that discourse is frequently interrupted in the (analytic) L2 classroom. While the results showed that metalinguistic feedback and elicitations lead to learner uptake

98% of the time, it is also noted that the occurrence of recasts was noticeably higher when the focus of the lesson was on meaning. Furthermore, successful uptake was higher when correction moves consisted of recasts or explicit correction. In an analytic classroom where meaningful interaction plays little to no role, these recasts resemble explicit error corrections and as such are more likely to be interpreted by the learner as feedback on form. This is reflected in the results of this thesis, where the highest overall mean stage of question development was produced by learners in the recast group. Thus, the results and analyses discussed here indicate that in a communicative classroom, providing recasts intensively in task-based interaction might prove to be an effective way of drawing learners' attention to form.

While the learners in the recast group outperformed the learners in the recast + prompt group in overall stage production, it is important to note that in terms of improvement and grammaticality the three treatment groups were relatively equal. It is also interesting that learners in the recast + prompt group were the only participants to produce a higher overall mean stage of questions on post-test 1. Further discussion of the recast + prompt group is provided in the following section which examines the results of post-test 2 in relation to the fourth and final research question concerning long-term effects of feedback. In particular, the evidence for *delayed* effects of recast + prompt is discussed.

5.4 Long-term Feedback Effects

The fourth research question enquired whether positive effects of either form of interactive feedback would be maintained over time. The research hypothesis associated

with this question predicted that any positive effects of either form of interactive feedback would be maintained over time. The results provide some support for this prediction; the recast + prompt group showed more improvement with each succeeding test. This final section of Chapter Five discusses the implications of the results and analyses pertaining to the delayed post-test; after providing a review of the results, the findings are discussed in relation to studies by Mackey (1999) and Nassaji (2007a) and provide support for their conclusions.

To review, the learners in the elicitation group produced the most developmentally advanced questions on post-test 2; no learners in any other group produced a question higher than Stage 4 on this delayed post-test. However, there was no statistically significant difference among any of the groups in terms of overall stage production; learners were each producing questions at approximately the same level of development. In addition, the elicitation group also showed the highest grammatical accuracy on post-test 2, although again there was no statistically significant difference among any of the four conditions. This finding lends support to Lyster's (2004) and Lyster and Ranta's (1997) studies that found elicitations and other forms of feedback to be more effective than recasts. Furthermore, the results relating to improvement of overall stage production revealed that all of the treatment groups showed more (although not statistically significant) improvement than the control group on post-test 2. This finding provides further support for positive effects of feedback in general.

Despite the initial difference found between the learners in the control and recast + prompt groups, the recast + prompt group showed the highest level of improvement of all groups. In fact, learners in this group improved more on the delayed post-test than on

the immediate one, suggesting that positive effects of recasts in combination with prompts may be delayed. This finding is supported by Mackey's (1999) study of input and interaction on L2 development. In her study, 34 adult ESL learners participated in task-based interaction and were observed for increases in development stages. The results found significant positive effects for interactionally modified input, but only in the delayed post-tests that were administered one week and one month after treatment. Mackey's results provide support for the finding of this research that at least one form of interactional feedback, namely recasts in combination with prompts, have delayed positive effects on question development. In addition, the learners in the elicitation group maintained the same level of improvement across both post-tests, suggesting long-term effects of elicitation. This finding is comparable to those found in studies by Mackey and Oliver (2002) and McDonough and Mackey (2006) such sustained effects were found for recasts. In contrast to these studies, however, there were no long-term or delayed effects found for learners in the recast group; the noticeably high levels of improvement observed by these learners on post-test 1 were not sustained for the second post-test. A discussion of Nassaji's (2007a, in press) studies of recasts and elicitation sheds some light on these results.

In his study of the immediate repair of erroneous utterances following recasts or elicitation, Nassaji (2007a) focused on the nature of each type of feedback and how it influenced learner repair. The study involved 42 adult intermediate ESL students in dyadic task-based interactions with two native English-speaking teachers. Both recasts and elicitation were found to increase the accuracy of immediate repair when they were paired with explicit prompts. Interestingly, recasts were found to be more effective than

elicitations when provided with a prompt that enhanced the recast in some way. This study substantiates the importance of saliency when providing interactional feedback. Furthermore, in a related study of elicitations and reformulations in dyadic interaction, Nassaji (in press) identified six subtypes of reformulations and five subtypes of elicitation, with the subtypes that included a prompt leading to a higher rate of successful student repair. Although the results of the immediate post-test employed in this thesis found that recasts were more effective alone than with a prompt, the results of the delayed post-test reflect Nassaji's findings and indicate the importance of longitudinal studies in determining the effects of interactional feedback.

In conclusion, while the provision of intensive recasts may not reflect naturalistic settings outside of the language classroom, the purpose of investigating feedback lies in the role of the L2 teacher. As noted by Lyster (1998), recasts are often ambiguous in meaning-oriented classrooms because teachers often reformulate perfectly acceptable utterances as well as erroneous ones. Research has shown that a purely communicative classroom where little or no attention to form occurs is not as effective as a classroom where learners' attention is drawn to noticing form. Such a focus on form may be provided by a teacher's consistent and intensive provision of recasts during tasks in which students are made aware of a particular form.

The tasks carried out in this study allowed the students, indeed required the students to respond in some fashion to the recasts, elicitations and recasts in combination with prompts. So, while previous studies may have found recasts to be less effective, when students are provided with *intensive* recasts in an environment where it is acceptable to modify their utterance (such as in an L2 classroom), recasts may actually be

more effective than previously suggested. In particular, the results of this study suggest that recasts are most effective when they are combined with a prompt. This likely increases the explicitness and/or saliency of the recast and increases the probability that it will be perceived as a focus on form. This study also provides support for both sustained and delayed effects of feedback; the enhanced saliency provided by the teacher in the recast + prompt group allowed the students to maintain and improve upon their levels of development.

Chapter Six

CONCLUSION

This chapter presents a conclusion to the research in four sections. The first section provides a summary of the goals, design and results of the study. The second section considers its limitations and is followed by a section on possible avenues of future research. The fourth and final section describes the possible contributions this study has made to the field of SLA research and the effectiveness of recasts and elicitations in the L2 classroom in particular.

6.1 Summary of Research

The purpose of this study was to investigate the effects of recast, elicitation, and recast + prompt on the development of question formation by ESL learners in a previously established classroom environment. The research was focused on four main questions: (1) Do learners who receive interactional feedback in the form of recasts and/or elicitation progress to higher stages of question development? (2) Do learners who receive recasts progress to a higher level of question development than learners who receive elicitation? (3) Is it more effective to provide recast and elicitation together than to provide only one of these types of feedback? (4) Are positive effects of either form of interactive feedback maintained over time?

In order to investigate these research questions, the study followed a pre-test/post-test design and was carried out over 7 weeks with four classes of intermediate L2 learners of English and two teachers. The project was intended to represent typical classroom activity while still controlling the type of feedback provided within each class. Each of

the 7 sessions involved dyadic task-based interaction between the teachers and their students. Students in each class were divided into two teams and took turns asking the teacher a question using a key word that was necessary for each team to complete their task. treatment took place once a week for 4 weeks, during which learners in the three treatment groups received a single form of interactional feedback (recast, elicitation or recast + prompt) in response to errors while learners in the control received no such feedback. L2 development was measured in two ways: First, the overall developmental stage of question production was compared across the pre-test, immediate post-test and delayed post-test. Second, the grammaticality of questions was compared across tests to determine any intra-stage development. This research was motivated by the mixed results observed in the literature on the effects of recasts and other types of interactional feedback as a means of drawing learners' attention to form in communicative-oriented classrooms. While the results of this study do not clearly show whether one type of feedback is more effective than another, they do indicate some short-term effects for recasts, sustained effects for elicitation, and delayed effects for recast + prompts.

The results provide support for the first research hypothesis that learners who received interactional feedback would produce questions at a higher stage than learners who did not receive any feedback. This was true of both measures of development; learners who received feedback produced higher stage and more grammatically accurate questions than learners in the control group. The results only partially support the second hypothesis that learners who received recasts would not progress to a higher level of question development than learners who received elicitation. While learners who received recasts produced the highest mean stage of questions on the immediate post-test,

learners who received elicitation produced more grammatical questions. These results mirror the mixed findings of previous studies; while some studies have found support for recasts (e.g., Han, 2002), others have found other types of feedback to be more effective (e.g., Lyster & Ranta, 1997). However, the results of this study also indicate that both recasts and elicitations may be effective when more than one measurement of development is used. The results of the immediate post-test do not support the third research hypothesis that learners who received both recasts and elicitation would produce questions at a higher stage than learners who received only one form of feedback. However, while learners in the recast group produced higher stage questions on post-test 1, learners in the recast + prompt group showed the highest level of improvement on post-test 2, thereby providing partial support for the fourth research hypothesis which predicted that positive effects of interactional feedback would be maintained over time.

6.2 Limitations

Despite the efforts made to address any potential problems during the design and data collection phases of this study, this project has some limitations. One obvious limitation of this study concerns the limited feedback provided to each group; as each type of interactional feedback has a unique conversational role and elicits different responses from students (Lyster & Mori, 2006), it follows that the most successful L2 classroom would be the one where a variety of feedback types is used. Thus, this study may not represent the L2 classroom as accurately as intended. While the purpose of this study was to examine three types of feedback individually, it does not provide any

insights into how and when teachers provide specific feedback moves in the course of interaction.

A related methodological imperfection is the highly repetitive nature of the tasks the students performed and the relatively short negotiations that occurred. Ideally, task-based interactions would involve lengthier negotiations for meaning that would allow more opportunity for learners to notice a teacher's feedback. In her examination of the role of recasts in interaction, Braidı found that recasts were more successful when provided within longer negotiations for meaning (Braidı, 2002). Unfortunately, the length of negotiations in this study was typically quite short.

In addition, only a limited amount of students' output was measured, so the absence of a higher stage question may not necessarily entail the student's inability to produce such a question. A study incorporating most or all of learners' production would be beneficial and more accurate. Finally, another shortcoming of this study regards the small number of participants in each group; this limitation makes it difficult to make any strong generalisations concerning the research questions.

6.3 Future Research

It is difficult to make direct comparisons between the results of this study and those that have evaluated feedback effects with only one way of measurement. However, it would be interesting to see how the results of this study compare to similar studies where the type of feedback is controlled but where there is no manipulation of tasks. Thus, one project for future investigation might involve a training session for teachers to

practice providing a single type of feedback throughout their regular scheduled class. Such a study would represent a typical L2 classroom, employ both measures of development, and still control the type of interactional feedback provided. In addition, this study focused on feedback provided during planned reactive FonF; future quasi-experimental studies should focus on incidental and preemptive FonF to determine the effectiveness of these feedback types in a various contexts.

Future studies might also replicate the present study with other slight modifications; for example, elements concerning students' background, first language, etc. may be controlled and/or analysed in order to test for possible effects. In particular, it would be interesting to examine the relationship between learners' previous classroom experiences and how they respond to different types of feedback. For example, a learner who is not accustomed to a communicative classroom may not notice or interpret implicit feedback as a sign that their utterance was incorrect. The design could also be made considerably longer with more frequent treatment sessions each week so that students had more time to get used to and respond to the implicit feedback provided. Another modification could include both beginner, intermediate, and advanced students in order to test the effects of interactional feedback on learners of various proficiencies. It is possible that learners in this study were at slightly different levels of development, so while the recast learners were ready to advance to a higher stage, the elicitation learners were only ready to advance within their particular stage. As well, the effects of *written* implicit negative feedback could be examined in a similar context. In addition, the effects of explicit versus implicit feedback, as well as positive versus negative evidence could be investigated with a similar design as the present study.

Finally, it is important that future research also attempt such hybrid projects in order to account for feedback effects as accurately as possible without intruding upon the typical interactions that naturally occur in the L2 classroom. Such hybrid or quasi-experimental research can only help to determine more conclusively the effects of different types of feedback.

6.4 Contributions

This study provides some support for the role of interactional feedback in L2 development in the L2 classroom. Specifically, the findings suggest that recasts and elicitations may both be effective in different ways. As this was the first study to measure development in terms of both developmental stages and grammatical accuracy, this study contributes to the growing body of literature concerning the similarities and differences in how different types of interactional feedback function in the L2 classroom. In addition, the design of this study allowed for specific types of feedback to be controlled in a relatively natural language learning environment. The participants' regular teacher provided the feedback in tasks which represented the current themes of the course.

A second contribution this study makes to the field of SLA regards the long-term effects of feedback. All of the students who received feedback showed more improvement on the second post-test than students in the control group. Furthermore, students in the recast + prompt group actually showed more improvement on the delayed post-test than on the immediate post-test. These results indicate that feedback may be

effective when provided consistently over time, and that individual feedback types should not necessarily be disregarded if they are not immediately observed to be effective.

Finally, this research provides some support for previous findings that recasts are more effective when their saliency is enhanced in some way (e.g., Lochtman, 2002; Nassaji, 2007a, 2007c). In this study, each type of feedback was provided intensively during a task that was explicitly focused on question production. Both of these findings have pedagogical implications; teachers of communicative classrooms should perhaps not be too cautious about providing more explicit negative feedback in their interactions with students. While the role *implicit* feedback provided during interaction is important, it seems that learners may often ignore the signals about the unacceptability of their utterance that occur during interaction. Thus, it is likely that enhancing feedback to make it more explicit, and therefore perhaps more noticeable, plays an important role in the L2 classroom.

REFERENCES

- Ammar, A., & Spada, N. (2006). One size fits all? Recasts, prompts, and L2 learning. *Studies in Second Language Acquisition*, 28, 543-574.
- Basturkmen, H., Loewen, S., & Ellis, R. (2004). Teachers' stated beliefs about incidental focus on form and their classroom practices. *Applied Linguistics*, 25, 243-272.
- Braidi, Susan M. (2002). Reexamining the role of recasts in native-speaker/nonnative-speaker interactions. *Language Learning*, 52, 1-42.
- Brown, H.D. (2000). *Principles of language learning and teaching*. New York: Addison Wesley Longman, Inc.
- Crookes, G., & Chaudron, C. (2001). Guidelines for language classroom instruction. In M. Celce-Murcia (Ed.) *Teaching English as a Second or Foreign Language* (3rd ed.). Boston: Heinle & Heinle, Thomson Learning, Inc.
- Doughty, C., & Varela, E. (1998). Communicative focus on form. In C. Doughty & J. Williams (Ed.) *Focus on Form in Classroom Second Language Acquisition* (pp. 114-138). New York: Cambridge University Press.
- Ellis, R. (2001). Introduction: Investigating form-focused instruction. *Language Learning*, 51, 1-46.
- Ellis, R., Basturkmen, H., & Loewen, S. (2001). Preemptive focus on form in the ESL classroom. *TESOL Quarterly*, 35, 407-432.
- Ellis, R., Basturkmen, H., & Loewen, S. (2002). Doing focus-on-form. *System*, 30, 419-432.
- Han, Zhaohong. (2002). A Study of the impact of recasts on tense consistency in L2 output. *TESOL Quarterly*, 36, 543-572.

- Havranek, G. (2002). When is corrective feedback most likely to succeed? *International Journal of Educational Research*, 37, 255-270.
- Hawkes, L. (2007). *Recasts revised: The role of recasts in error detection and correction by adult ESL students*. Unpublished master's thesis, University of Victoria, Victoria, British Columbia, Canada.
- Instructors' Handbook*. (2006) English Language Centre, University of Victoria, BC.
- Leeman, Jennifer. (2003). Recasts and second language development. *Studies in Second Language Acquisition*, 25, 37-63.
- Lightbown, P. (1998). The importance of timing in focus on form. . In C. Doughty & J. Williams (Ed.) *Focus on Form in Classroom Second Language Acquisition* (pp. 177-196). New York: Cambridge University Press.
- 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, 271-283.
- Loewen, S. & Philp, J. (2006). Recasts in the adult English L2 classroom: Characteristics, explicitness, and effectiveness. *The Modern Language Journal*, 90, 536-556.
- Long, M.H., & Robinson, P. (1998). Focus on form: theory, research, and practice. In C. Doughty & J. Williams (Ed.) *Focus on Form in Classroom Second Language Acquisition* (pp. 15- 41). New York: Cambridge University Press.
- Lyster, R. (1998a). Negotiation of form, recasts, and explicit correction in relation to error types and learner repair in immersion classrooms. *Language Learning*, 48, 183-218.

- Lyster, R. (1998b). Recasts, repetition, and ambiguity in L2 classroom discourse. *Studies in Second Language Acquisition*, 20, 51-81.
- Lyster, R. (2002). Negotiation in immersion teacher-student interaction. *International Journal of Educational Research*, 37, 237-253.
- Lyster, R. (2004). Differential effects of prompts and recasts in form-focused instruction. *Studies in Second Language Acquisition*, 26, 399-432.
- Lyster, R. & Mori, H. (2006). Interactional feedback and instructional counterbalance. *Studies in Second Language Acquisition*, 28, 269-300.
- Lyster, R. & Ranta, L. (1997). Corrective feedback and learner uptake. *Studies in Second Language Acquisition*, 20, 37-66.
- Mackey, A. (1999). Input, interaction, and second language development: An empirical study of question formation in ESL. *Studies in Second Language Acquisition*, 21, 557-587.
- Mackey, A., Al-Khalil, M., Atanassova, G., Hama, M., Logan-Terry, A. and Nakatsukasa, K. (2007). Teachers' intentions and learners' perceptions about corrective feedback in the L2 classroom. *Innovation in Language Learning and Teaching*, 1, 129-152.
- Mackey, A., Gass, S. & McDonough, K. (2000). How do learners perceive interactional feedback? *Studies in Second Language Acquisition*, 22, 471-497.
- Mackey, A. & Oliver, R. (2002). Interactional feedback and children's L2 development. *System*, 30, 459-477.
- Mackey, A. & Philp, J. (1998). Conversational interaction and second language development: recasts, responses, and red herrings?. *The Modern Language*

- Journal*, 82, 338-356.
- McDonough, K., & Mackey, A. (2006). Responses to recasts: Repetitions, primed production, and linguistics development. *Language Learning*, 56, 693-720.
- Nabei, T., & Swain, M. (2001). Learner awareness of recasts in classroom interaction: A case study of an adult EFL student's second language learning. *Language Awareness*, 11, 43-63.
- Nassaji, H. (2007a). Elicitation and reformulation and their relationship with learner repair in dyadic interaction. *Language Learning*, 57, 511-548.
- Nassaji, H. (2007b). "Reactive focus on form through negotiation on learners written errors." In S. Fotos & H. Nassaji (Eds.). *Form-focused instruction and teacher education: Studies in honour of Rod Ellis* (pp. 117-129). Oxford: Oxford University Press.
- Nassaji, H. (2007c). "Focus on form through recasts in dyadic student-teacher interaction: A case for recast enhancement. In C. Gascoigne (Ed.), *Assessing the impact of input enhancement in second Language Education* (pp. 53-69). Stillwater, OK: New Forums Press.
- Nassaji, H. (in press). The effects of recasts and elicitations in dyadic interaction and the role of feedback explicitness. *Language Learning*, 59.
- Nicholas, H., Lightbown, P., & Spada, N. (2001). Recasts as feedback to language learners. *Language Learning*, 51, 719-758.
- Ohta, Amy Snyder. (2000). "Rethinking recasts: A learner-centered examination of corrective feedback in the Japanese classroom" in J. K. Hall and L. Verplaeste

- (eds.): *The construction of second and foreign language through classroom interaction*. Mahwah, NJ: Erlbaum.
- Panova, I. & Lyster, R. (2002). Patterns of corrective feedback and uptake in an adult ESL classroom. *TESOL Quarterly*, 36, 573-595.
- Pienemann, M., & Johnston, M. (1987). Factors influencing the development of language proficiency. In D. Nunan (Ed.), *Applying second language acquisition research* (pp. 45-141) Adelaide, Australia: National Curriculum Resource Centre, AMEP.
- Pienemann, M., Johnston, M., & Brindley, G. (1988). Constructing an acquisition-based procedure for second language assessment. *Studies in Second Language Acquisition*, 10, 217-243.
- Poole, A. (2005). The kinds of forms learners attend to during focus on form instruction: a description of an advanced ESL writing class. *Asian EFL Journal*, 7, 75-87.
- Sharwood Smith, M. (1993). Input enhancement in instructed SLA: Theoretical bases. *Studies in Second Language Acquisition*, 15, 165-179.
- Sheen, Y. (2004). Corrective feedback and learner uptake in communicative classrooms across instructional settings. *Language Teaching Research*, 8, 263-300.
- Spada, N., & Lightbown, P. (1999). Instruction, first language influence, and developmental readiness in second language acquisition. *The Modern Language Journal*, 83, 1-22.
- Swain, M. (1995). Three functions of output in second language learning. In G. Cook, & B. Siedlhofer (Eds.), *Principle and Practice in Applied Linguistics* (pp.125-144). Oxford, UK: Oxford University Press.

- Swain, M. (2000). The output hypothesis and beyond: Mediating acquisition through collaborative dialogue. In J. Lantolf (Ed.), *Sociocultural theory and second language learning* (pp. 97-114). Oxford, UK: Oxford University Press.
- VanPatten, B. (1996). *Input processing and grammar instruction: Theory and research*. Norwood, NJ: Ablex.
- Williams, J. (1999). Learner-generated attention to form. *Language Learning*, 49, 583-625.
- Wray, A. (2002). *Formulaic Language and the Lexicon*. Cambridge, UK: Cambridge University Press.

Appendix A

SLEP Test: Sample Questions

The first section of the SLEP Test measures students' ability to understand spoken English. Example 1 provides a sample question from the first section of the test.¹² The students must match one of four recorded sentences with a picture in the test book.

Example 1

On tape: *Look at the picture marked number 1.*



On tape:

- (A) *The boy is typing on the keyboard.*
- (B) *No one is looking at the computer.*
- (C) *One of the girls is pointing to the screen.*
- (D) *They're putting a computer into a box.*

The second part of the SLEP Test measures students' ability to understand written English. Example 2 provides a sample question where students must complete passages by selecting the appropriate words or phrases from a set of four choices.

Example 2

Most animals

(A) live
(B) they live
(C) are living
(D) that live

 in the desert do not take in water from open sources.

¹² Examples from the SLEP Test Manual retrieved September 25, 2007, from, <http://www.ets.org>. Examples have been modified to fit the document format.

Appendix B

Sample Crossword

TEAM BLUE

	¹ H							² U							
	A					³		T				⁴		⁵ R	
	⁶ R	E	P	U	T	A	T	I	O	N		⁷		E	
	M							L						L	
					⁸			I						⁹ A	D
¹⁰ B								T						X	
E		¹¹			¹²			I						E	
N								E						D	
¹³ E	N	T	H	U	S	I	¹⁴ A	S	T	¹⁵ I	C				
F															
I															
T									¹⁶ J	E	R	K			
					¹⁷ F	U	N	D							
					¹⁸										

ACROSS

7. Oh no, I forgot my purse. Could you wait _____ while I run and get it?
12. He was so _____ it was almost Christmas - he really needed to take a break!
18. When she first went to Sweden, she felt _____ a lot of the time because everything was so different.

DOWN

3. Do you think it's a good idea to _____ smoking from outside patios as well as restaurants?
4. Whew! I'm going straight to bed, I'm _____.
8. He had never been so _____ as when he had to write three final exams in one day.
11. Only children who both study and play will reach their full _____.
14. Britain is an _____ of Canada.
15. Jake was so shy he had a friend ask Johanna if she was _____ in going out with him.

Appendix C

Sample Story Completion Task

Team Blue: The Story of My Grandparents

My grandfather was a veteran of World War II and he often told me stories about his experiences during the war. My favourite story was how he met my grandmother. He was in a small European town and saw a beautiful woman _____ on a porch swing, but when he turned to _____ at her he fell on some ice and broke his leg! The mysterious woman _____ to him - she was a nurse and tried to carry him to the nearest hospital. He couldn't even _____ up, and he certainly couldn't walk to the hospital himself. Another soldier _____ and took him to the hospital – and the beautiful woman went with them. Grandfather said the woman held his hand while the doctor put his leg in a cast. He had to stay _____ in bed for two weeks, but he got to see the pretty nurse every day. She would tell him stories of swimming in a clear blue lake when she was young, of _____ off high rocks into the water, and of catching fireflies with her hands. Grandfather said that these stories were the best thing anyone had ever _____ him. Finally, he climbed out of bed, _____ on one knee, and asked the nurse to marry him! Of course, the nurse was my grandmother – they were married for over fifty years.

Team Yellow: The Story of My Grandparents

My grandfather was a veteran of World War II and he often told me stories about his experiences during the war. My favourite story was how he met my grandmother. He was in a small European town and saw a beautiful woman sitting on a porch swing, but when he _____ to wave at her he _____ on some ice and broke his leg! The mysterious woman ran to him - she was a nurse and tried to _____ him to the nearest hospital. He couldn't even stand up, and he certainly couldn't _____ to the hospital himself. Another soldier picked him up and _____ he to the hospital – and the beautiful woman went with them. Grandfather said the woman _____ his hand while the doctor put his leg in a cast. He had to stay lying down in bed for two weeks, but he got to see the pretty nurse every day. She would tell him stories of _____ in a clear blue lake when she was young, of jumping off high rocks into the water, and of c _____ fireflies with her hands. Grandfather said that these stories were the best thing anyone had ever given him. Finally, he _____ out of bed, knelt on one knee, and asked the nurse to marry him! Of course, the nurse was my grandmother – they were married for over fifty years.

Appendix D

Background Information

1. Name _____
2. Gender: Male Female
3. Age: _____
4. Language Level: _____
5. First Language: _____
6. How long have you been in Canada? _____
7. At what age did you start to learn English? _____
8. How long have you been studying English?

What level of education have you completed? (For example, high school/university degree, etc.) _____