Studies in Second Language Learning and Teaching
Department of English Studies, Faculty of Pedagogy and Fine Arts, Adam Mickiewicz University, Kalisz
SSLLT 6 (1). 2016. 19-41
doi: 10.14746/ssllt.2016.6.1.2
http://www.ssllt.amu.edu.pl

The role of extensive recasts in error detection and correction by adult ESL students

Laura Hawkes
British Columbia Ministry of Education, Victoria, BC, Canada
laura.hawkes@gov.bc.ca

Hossein Nassaji
University of Victoria, Victoria, BC, Canada
nassaji@uvic.ca

Abstract

Most of the laboratory studies on recasts have examined the role of intensive recasts provided repeatedly on the same target structure. This is different from the original definition of recasts as the reformulation of learner errors as they occur naturally and spontaneously in the course of communicative interaction. Using a within-group research design and a new testing methodology (video-based stimulated correction posttest), this laboratory study examined whether extensive and spontaneous recasts provided during small-group work were beneficial to adult L2 learners. Participants were 26 ESL learners, who were divided into seven small groups (3-5 students per group), and each group participated in an oral activity with a teacher. During the activity, the students received incidental and extensive recasts to half of their errors; the other half of their errors received no feedback. Students' ability to detect and correct their errors in the three types of episodes was assessed using two types of tests: a stimulated correction test (a video-based computer test) and a written test. Students' reaction time on the error detection portion of the stimulated correction task was also measured. The results showed that students were able to detect more errors in error+recast (error followed by the provision of a recast) episodes than in error-recast (error and no recast provided) episodes (though this difference did not reach statistical significance).

They were also able to successfully and partially successfully correct more errors in error+recast episodes than in error-recast episodes, and this difference was statistically significant on the written test. The reaction time results also point towards a benefit from recasts, as students were able to complete the task (slightly) more quickly for error+recast episodes than for error-recast episodes.

Keywords: error detection; error correction; recast; corrective feedback; grammaticality

1. Introduction and literature review

Within the field of second language acquisition (SLA) research, *corrective feedback,* a response by a teacher or other interlocutor that attempts to signal to a nonnative speaker (NNS) the incorrectness/ungrammaticality of the NNS's utterance, has received a great deal of attention in recent years. Among corrective feedback types, recasts have received the most attention from researchers. Early observation classroom studies have found that recasts occur more often than any other type of corrective feedback in natural L2 classrooms; this observation was found in ESL classrooms (Havranek, 1999; Panova & Lyster, 2002), French immersion classrooms (Lyster, 1998; Lyster & Ranta, 1997), a German L2 classroom (Lochtman, 2002), and in NS-NNS dyadic interaction (Iwashita, 2003; Nassaji, 2007, 2009). Recasts are generally considered to be a form of implicit correction in that they do not interrupt the flow of meaningful interaction (see Doughty & Varela, 1998; Goo & Mackey, 2013; Nicholas, Lightbown, & Spada, 2001), and teachers may use recasts because they do not slow down the flow of communication as much as some other types of corrective feedback.

Theoretically, recasts have been assumed to be beneficial for language learning because they provide positive evidence: They supply the learner with the correct form (Ellis & Sheen, 2006; Nicholas et al., 2001). It has also been claimed that recasts provide negative evidence, which has been considered necessary for adults L2 learners. Recasts may also increase the perceptual saliency of the target form as the juxtaposition of the learner's utterance and the teacher's recast highlight the error (Farrar, 1990; Saxton, 1997). Finally, it has been proposed that recasts promote interaction. Long's interaction hypothesis states that in addition to input, participation in interaction is needed for second language learning to occur (Long, 1981, 1983, 1996). Closely related to the idea of interaction is Swain's (1985) concept of pushed output, which proposes that learners must be pushed to produce modified output in order for L2 learning to occur; recasts are proposed to push learners to change their output.

Given the theoretical benefits for recasts, many studies have examined the beneficial effects of recasts (Ammar & Spada, 2006; Dilans, 2010; Ellis, Loewen, & Erlam, 2006; Iwashita, 2003; Loewen & Philp, 2006; Lyster, 2004; Lyster & Ranta, 1997; Lyster, Saito, & Sato, 2012; Mackey & Philp, 1998; Nabei & Swain, 2002; Nasssaji, 2007, 2009; Panova & Lyster, 2002; Saito, 2013 among others). These studies, the results of which have also been summarised in a number of recent reviews and meta-analyses (e.g., Li, 2010; Lyster & Saito, 2010; Lyster, Saito & Sato, 2012; Nassaji, 2015; Sheen, 2011), have reported positive effects for recasts in general. However, most of such studies have provided recasts intensively and repeatedly on the same target structure, which can be different from the definition of recasts as reformulation of the learner errors as they occur incidentally in the course of interaction. Of course, some early studies on recasts were conducted in natural second language classrooms and were observational in nature (Ellis, Basturkmen, & Loewen, 2001; Lochtman, 2002; Lyster, 1998; Lyster & Ranta, 1997; Panova & Lyster, 2002). However, these studies used *uptake*, with uptake being defined as a learner's production of the correct form following the corrective feedback, to see how effective the various types of corrective feedback were. The results of these studies showed that recasts might not be as effective as other types of feedback as they tended to produce less uptake than, for example, clarification requests. However, despite such studies, a number of researchers have questioned whether uptake and other measures of noticing can be used as valid measures of the beneficial role of recasts (see Loewen, 2005; Loewen & Philp, 2006; Mackey & Philp, 1998; Nicholas et al., 2001). The use of uptake is perhaps most often linked to Schmidt's noticing hypothesis (Schmidt, 1990, 1995), which proposed that attention is necessary for language learning to occur. However, as Mackey and Philp (1998) point out, while uptake can signal the presence of noticing, no uptake does not necessarily indicate the absence of noticing.

In the movement away from uptake as the sole measure of the effectiveness of recasts, a number of researchers turned to pre-/posttest designs. With the introduction of pre-/posttests also came the increased use of preselected target linguistic forms. The need for such targets is clear: It is nearly impossible to pretest linguistic forms when you do not know on which forms the learner will produce errors and receive spontaneous recasts. Table 1 summarizes a number of pre-/posttest studies that have examined recasts.

Table 1 Summary of a sample of pre-/posttest recast studies

Study	Measure of L2 learning	Feedback types examined	Context	Preselected target form	Intensive/ extensive
Mackey & Philp (1998)	Pre-/posttests	Recasts	Dyads	English questions	Intensive
McDonough & Mackey (2006)	Pre-/posttests and uptake	Recasts	Dyads	English questions	Intensive
Iwashita (2003)	Pre-/posttests	Naturally arising*	Dyads	Japansese locative-initial construction & *te*-verb	Extensive
Leeman (2003)	Pre-/posttests	Recasts vs. negative evidence vs. enhanced saliency	Dyads	Spanish noun-adjective agreement	Intensive
Han (2002)	Pre-/posttests	Recasts	Small group	Tense consistency	Intensive
Lyster (2004)	Pre-/posttests	Recasts vs. prompts	Classroom	French noun gender	Intensive
Ammar & Spada (2006)	Pre-/posttests	Recasts vs. prompts	Classroom	3rd person singular determiners	Intensive
Ellis, Loewen, & Erlam (2006)	Pre-/posttests	Recasts vs. metalinguistic feedback	Classroom	Past tense *–ed*	Intensive
Ishida (2004)	Pre-/posttests	Recasts	Dyads	Japanese *–te i-*(*ru*)	Extensive
Nassaji (2006)	Pre-/posttests	Recasts vs. elicitations	Dyads		Extensive
Loewen & Philp (2006)	Posttests and uptake	Naturally arising,* with a focus on reporting recasts	Classroom		Extensive

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

While the above studies have found some benefits for recasts, most of them employed intensive recasts directed at preselected target linguistic forms. Therefore, it is not clear whether their findings were due to the recasts themselves or the intensity of the feedback. A few of those studies have examined the effects of extensive recasts. Ishida (2004) attempted to lessen the unnatural situation of providing intensive recasts by providing recasts "whenever [the researcher] felt the need to confirm the meaning of the message" (p. 340). While caution should be employed due to the small number of students involved (*N* = 4), the results of Ishida's study showed a positive correlation between the number of recasts a learner received and their accurate use of *–te i-*(*ru*). Nassaji (2006) went one step further than Ishida by examining truly extensive recasts using an innovative pre-/posttest design with no preselected target form. Each of 42 learners was asked to write a description based on a series of pictures; the descriptions were then collected and the NNS was asked to orally describe the story. During this oral interaction, corrective feedback was provided whenever the NS felt it was appropriate. After the oral interaction, the written story was returned to the student and he/she was asked to make corrections. A similar delayed posttest was given two weeks later. Results showed that learners successfully corrected more of the errors that had received recasts than the errors that had received elicitations. Loewen (2005) conducted a classroom study on

the effectiveness of spontaneous focus-on-form (both reactive and preemptive) on L2 learning. Although his study did not single out recasts for examination, the data was reexamined in Loewen and Philp (2006) with a close focus on recasts. Individualized immediate and delayed posttests were created for each student based on focus-on-form episodes (FFEs) that had occurred in the classroom interaction. The posttests consisted of oral suppliance, correction, and pronunciation tasks. As the feedback was spontaneous and the researchers did not know ahead of time which forms would receive feedback, no pretests could be administered. Instead, the initial errors made by the students served as a type of pretest (i.e., if the student made the error, it indicated that their knowledge of that form was to some degree incomplete).

The above findings suggest that recasts, even when provided incidentally and extensively, may be beneficial to L2 students, both in dyadic interaction and in classroom interaction. However, more research is needed in this area, especially in the area of nondyadic interaction. In addition, there are some methodological concerns with these studies. First, there is concern over the naturalness of dyadic studies. In dyadic studies, learners receive undivided individual attention from a NS over an extended period of time, which is not the norm in a typical L2 classroom. As such, it is possible that students pay more attention in dyadic interaction. Ellis and Sheen (2006), Lyster (2004), and Nicholas, Lightbown, and Spada (2001) have pointed out that dyadic interaction may produce different results than larger-group interaction. Second, there is the concern that studies that use preselected target linguistic forms and/or intensive recasts may increase learners' attention to recasts and the forms targeted by the recasts more than natural L2 classrooms (see Ellis & Sheen, 2006; Nicholas et al., 2001). Since the pretest, tasks during treatment sessions, and posttests were all designed to elicit the target form, the students may have picked up on this and directed their attention to that form. Third, there is the concern that studies that examine the effectiveness of providing intensive recasts in response to errors in the target form while ignoring all other types of errors may have little value when discussing the effectiveness of recasts in real L2 classrooms, which are generally provided extensively. Examples of recasts from observational studies have shown that recasts provided in L2 classrooms are generally provided in response to a wide range of errors (Loewen & Philp, 2006; Nassaji & Hawkes, 2006; Sabbagh, 1998; Sheen, 2006). Thus, as Ellis and Sheen (2006) pointed out, "the claim that recasts are most effective when they are focused and intensive (i.e., directed repeatedly at a single linguistic feature) is of little practical significance to teachers" (p. 597).

An additional problem that needs to be overcome in the use of posttests when examining spontaneous, extensive recasts is the matter of to what the

results of the posttest should be compared. Since it is not possible to give a pretest when you do not know what forms students will receive recasts on, how can we know whether the accuracy on posttests reveals something? We need to find something meaningful with which to compare it before we can make any claims about the beneficial role of recasts.

The above review has demonstrated that there are unresolved issues with previous studies surrounding varying results and methodological concerns. The concerns with previous recast studies are neither few nor trivial, and they make a strong call for further study of recasts, and in particular for new methodology in the study of extensive recasts. The purpose of the present study is to examine the effectiveness of spontaneous, extensive recasts in small-group interaction by employing a new testing methodology: *stimulated correction*. The *stimulated correction* testing method was designed to have the following advantages over previous measures of the effectiveness of recasts:

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

The study addressed the following research questions:

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

2. Method

The current study employed an experimental within-subject research design. Small groups of adult ESL students participated in an oral task with a teacher (the researcher) in which some of each student's errors received recasts and some of their errors received no feedback (were ignored).

2.1. Participants

In total, 26 adult ESL students participated in the study; three of the students did not complete day two of the study and their data were excluded from the analysis. The students were all enrolled in intermediate-level classes at the English Language Centre at the University of Victoria, Canada at the time of the study. The 23 students who completed the study had a mean age of 25.2 years, had been living in Canada for an average of 4.2 months at the time of the study, and came from seven different L1 backgrounds. Students were assigned to one of seven small groups on a first-come, first-serve basis, and the number of students in each group ranged from three to five students. The use of small groups in the present study was designed to mimic the way in which a small group would operate within a classroom. In this way, the present study could avoid the individualized attention that students receive in dyadic studies.

2.2. Procedure

Data collection took place over two days for each group. On day one, the small group of students met with the researcher in a small classroom. The students filled out a background questionnaire and then participated in an oral small-group activity, which was captured using a digital video camera and an external microphone. The small-group activity consisted of one task involving two parts: jigsaw (see Crookes & Gass, 1993; Johnson, 1981) and decision making (see Crookes & Gass, 1993; Doughty & Pica, 1986). The average time the groups took to complete the task was 38 minutes. During the activity, the researcher took on the role of the teacher and provided recasts following roughly half of each student's erroneous utterances. The remaining errors received no feedback (were ignored). In total, there were 141 erroneous utterances that received recasts and 150 erroneous utterances that received no feedback, leading to a ratio of 1.06 to 1.00.

The recasts in the present study were provided extensively. The recasts provided during the interaction had the following characteristics: (a) immediately followed a student's erroneous utterance, (b) repeated all or part of the student's utterance while reformulating the error(s), (c) did not change or add any information to the learner's target-like portion of the utterance, and (d) employed a rising intonation.

Some of these characteristics are illustrated in the following example, which uses (as all the other examples) the transcription conventions provided in the appendix:

Example 1

S1: They should look the your eye.
T: Look you in the eye?
S1: In the eye.

Each student was tested both orally (stimulated correction task) and in written form (written test). In order to create the tests, three types of episodes were identified in the video recordings: *error+recast, error-recast,* and *correct.* Error+recast episodes involved exchanges in which a student made an error and then received a recast; in error-recast episodes, the student produced an erroneous utterance but did not receive any feedback; correct episodes were those in which the student produced a grammatically correct utterance.

In total, 402 episodes were identified. 333 of these episodes (111 of each type) were randomly selected to serve as testing episodes and 69 episodes (3 per student) were selected to serve as practice episodes. The episodes were then edited using iMovie to create the short video clips, as in Examples 2, 3 and 4 below. The clips were edited to end with a student's utterance, and the student was required to judge the grammaticality of their final utterance in the stimulated correction task. In Examples 2, 3 and 4, the boxes around the utterances indicate which utterances were included in the edited clips; the symbol ♬ indicates an auditory cue, a "beep" that was inserted to direct the student's attention to the utterance of which they would be required to judge the grammaticality.

Example 2: An error+recast episode

T: Well, what do you guys think?
S1: I think if, if, uh Lisa can speak another language, she understands another student things ♬ so she can teach very well that country students.
T: Students from that country?
S1: Yeah.

Example 3: An error-recast episode

S2: Yeah, I think so.

T: You all agree! For the same reasons or different ones?
S1: Uh, I think uh Philip is more academic than Lisa, but he has enough- he te-
♬ he didn't have enough experience to teach to the student.
T: Mm hmm.
S1: So just uh he tea- he taught reading and writing-

Example 4: A correct episode

T: Mm hmm. Okay. So whose work experience do you like the best? Do you think?
S4: I think Lisa.
T: Okay, why?
S4: ♬ *Because she has uh lots of experience teaching ESL.*
Because three years she teach English and Lisa also uh it is relative in researching.
T: Mm hmm.

The day after the small-group interaction, each student completed first the stimulated correction task and then the written test individually in a quiet room. For the stimulated correction task, which was presented using SuperLab 4.0, the student was seated in front of a computer with headphones and a microphone. The procedure was as follows: (1) a screen with the words "The video will begin in 3 seconds," which was displayed for three seconds, (2) the playing of a video clip, (3) a screen with the words "Press ERROR or NO ERROR," which was displayed until the student pressed either the "ERROR" key or the "NO ERROR" key on the keyboard, which students were instructed to do as quickly as possible, and optionally (4) a screen with the words "Now correct it," which was displayed for 8 seconds and prompted the student to orally correct their error(s) (Screen 4 would appear only if the student had pressed the error key). Each student was tested only on those clips which involved their own utterances. Each student viewed an average of 14.5 video clips (the range was from 7 to 21). Reaction time was measured between the end of the video clip and the time at which the student pressed the error or the no-error key.

A written test was also created for each student. Each test question consisted of one written sentence (the utterance after the ♬ in the video clips). Students were instructed to read each sentence, indicate if there were any errors, and correct any errors. Sentences were presented in random order and students were allowed as much time as they needed to complete the written test.

3. Data analysis

The coding of the error correction responses involved comparing the student's utterance in the interaction with the student's modification of their utterance during the stimulated correction task and the written test. Students' modifications were coded as *successful modification* (the student corrected all the error(s)), *partially successful modification* (the student corrected only one/some of the error(s)), *unsuccessful modification* (the student failed to correct any errors), or *no modification* (no attempt to correct). For correct episodes only, the coding *successful* meant that the student had changed the (already correct) utterance

in such a way that it could also be considered grammatically correct and *unsuccessful* meant the student changed the correct utterance into an incorrect one.

Student modification of error+recast episodes were further coded to allow a more detailed analysis of how recasts may have influenced students' ability to modify their errors: *Successful* modifications were subcoded as either *successful/same as recast* or *successful/different than recast*. *Partially successful* modifications were subcoded as either *partial/same as recast* or *partial/different than recast*.

4. Results

4.1. Error detection results

As shown in Table 2, on the error detection task students indicated there were errors in error+recast and error-recast episodes significantly more often than in correct episodes on both the stimulated correction task and the written test, which was expected. Encouragingly, students detected the majority of the errors in their speech. On both the stimulated correction task and the written test, there was a trend for students to detect more errors on the error+recast episodes than on the error-recast episodes. On the stimulated correction task, students detected 76.6% of the errors in the error+recast episodes versus 71.2% in the error-recast episodes. However, chi-square analysis showed that the difference did not reach statistical significance [χ^2 (1, N = 222) = .840, p =.359]. On the written test, students detected 83.8% of the errors on the error+recast episodes and 82.0% of the errors on the error-recast episodes. A chi-square analysis found that the difference was not significant [χ^2 (1, N = 222) = .127, p = .722].

Table 2 Students' error detection responses

	Episode type	Errors detected		Errors not detected	
		n	%	*n*	%
Stimulated correction task	Error+recast	85	76.6	26	23.4
	Error-recast	79	71.2	32	28.8
	Correct	53	47.7	58	52.3
	Total	*217*	*65.2*	*116*	*34.8*
				χ^2 (2, N = 333) = 15.10, p = .001	
Written test	Error+recast	93	83.8	18	16.2
	Error-recast	91	82.0	20	18.0
	Correct	59	53.2	52	46.8
	Total	*184*	*82.9*	*38*	*17.1*
				χ^2 (2, N = 333) = 46.636, p = .000	

4.2. Error correction results: Stimulated correction task

As shown in Table 3, when the modification patterns for error+recast and error-recast episodes on the stimulated correction task were compared, the results showed a trend for a greater percentage of successful modifications on error+recast episodes (25.9%) than of error-recast episodes (17.7%); the same trend, but to a lesser degree, was found for partially successful modifications. Students also made more modification attempts for error+recast episodes (85.9%) than error-recast episodes (78.5%), suggesting that the recasts may have assisted them with some of their modifications. Although the differences in modification patterns in Table 3 show a benefit from recasts, a chi-square analysis found that this difference between error+recast and error-recast episodes was not significant [χ^2 (3, N = 164) = 2.681, p = .443].

Table 3 Error modification patterns (stimulated correction task)

	Total	Successful		Partial		Unsuccessful		Not modified	
		n	%	*n*	%	*n*	%	*n*	%
Error+recast	85	22	25.9	19	22.4	32	37.6	12	14.1
Error-recast	79	14	17.7	16	20.3	32	40.5	17	21.5
Correct	53	21	39.6	NA	NA	21	39.6	11	20.8
Total	217	57	26.3	35	16.1	85	39.2	40	18.4
								χ^2 (6, N = 217) = 18.73, p = .005	

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

Students' successful and partially successful modifications for error+recast episodes were also examined to see if the utterances were modified in the same way as the teacher-provided recast had modified their utterance. As Table 4 shows, when students successfully modified their errors, they were much more likely to do so in the same way as the recast had done (81.9%) than in a different way than the recast had done (18.1%). This difference was found to be significant (χ^2 [1, N = 41] = 4.011, p = .045).

Table 4 Same as recast vs. different than recast successful and partially successful modifications on error+recast episodes (stimulated correction task)

	Same as recast		Different than recast	
	N	%	*N*	%
Successful	18	81.9	4	18.1
Partially successful	10	52.6	9	47.4
Total	28	68.3	13	31.7
			χ^2 (1, N = 41) = 4.011, p = .045	

4.3. Error correction results: Written test

The error modification results for the written test are presented in Table 5. When the results of the three types of episodes were compared, a chi-square found a significant difference [χ^2 (6, N = 333) = 38.415, p = .000]. The most noticeable difference between the modification patterns was found in successful modifications, with correct episodes resulting in significantly more successful modifications than error+recast and error-recast episodes.

Table 5 Error modification patterns (written test)

	Total	Successful		Partial		Unsuccessful		Not modified	
		n	%	*n*	%	*n*	%	3	3.3
Error+recast	93	34	36.6	33	35.5	23	24.7	3	3.2
Error-recast	91	21	23.1	29	31.9	38	41.8	1	1.7
Correct	59	36	61.0	NA	NA	22	37.3	7	2.9
Total	243	91	37.4	62	25.5	83	34.2	3	3.3
								χ^2 (6, N = 333) = 38.415, p = .000	

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

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

As with the results from the stimulated correction task, the modification patterns of error+recast and error-recast episodes on the written test were examined in more detail. Table 6 presents the results of combining successful with partially successful modifications and unsuccessful with not modified responses. As can be seen, error+recast episodes resulted in many more successful and partially successful modifications than error-recast episodes (72.0% vs. 54.9%), and this difference was found to be significant [χ^2 (1, N = 222) = 5.807, p = .016].

Table 6 Combined error modification patterns, error+recast and error-recast episodes (written test)

	Successful & partial		Unsuccessful & not modified	
	n	%	*n*	%
Error+recast	67	72.0	26	28.0
Error-recast	50	54.9	41	45.1
Total	117	63.6	67	36.4
			χ^2 (1, *N* = 222) = 5.807, *p* = .016	

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

Student modifications on error+recast episodes were also examined in greater detail. Table 7 displays the frequency of students' successful and partially successful modifications according to whether they were done in the same way as the recast which had been provided or in a different way than the recast which had been provided. It was found that for both successful and partially successful modifications, students were more likely to modify their utterance in the same way as in the recast (82.4% of successful modifications and 66.7% of partially successful modifications). However, a chi-square showed that these differences did not reach significance [χ^2 (1, *N* = 67) = 2.176, *p* = .140].

Table 7 Same as recast vs. different than recast successful and partially successful modifications on error+recast episodes (written test)

	Same as recast		Different than recast	
	n	%	*n*	%
Successful	28	82.4	6	17.6
Partially successful	22	66.7	11	33.3
Total	50	74.6	17	25.4
			χ^2 (1, *N* = 67) = 2.176, *p* = .140	

Thus, in response to the question of whether errors that received recasts would be accurately corrected by students more often than errors that did not receive recasts, the results of the written test showed that students' modifications of their errors on error+recast and error-recast episodes were significantly different. Specifically, when successful was combined with partially successful and unsuccessful was combined with not modified, it was found that error+recast episodes lead to significantly more successful and partially successful modifications than error-recast episodes.

4.4. Comparison between stimulated correction and written test

Table 8 shows the differences in error modification between the stimulated correction task and the written test. Positive numbers indicate the students had more of that type of modification on the written test than on the stimulated correction task; negative numbers indicate that the students had fewer of that type of modification on the written test than on the stimulated correction task.

Table 8 Gains/losses in modifications from stimulated correction task to written test

	Successful		Partial		Unsuccessful		Not modified	
	n	%	*n*	%	*n*	%	*n*	%
Error+recast	+12	+10.7	+14	+13.1	-9	-12.9	-9	-10.8
Error-recast	+7	+5.4	+13	+11.6	+6	+1.3	-14	-18.3
Correct	+15	+21.4	NA	NA	+1	-2.3	-10	-19.1

In general, we see a trend for students to perform more favourably on the written test than on the stimulated correction task. However, the effect was not even across the three episode types. For example, while error+recast episodes had 12.9% fewer unsuccessful answers on the written test than on the stimulated correction task, for error-recast and correct episodes the number of unsuccessful modifications actually increased on the written test (+1.3% and +2.3% respectively).

4.5. Reaction time results

Reaction time (i.e., time between the end of the video and when the student pressed the error or the no-error key) results, which are presented in Table 9, show that students were fastest on the correct episodes (2.16076 sec.); students were slightly faster on the error+recast episodes (2.24956 sec.) than on the error-recast episodes (2.28148 sec.). However, a one-way ANOVA found that the differences in reaction times were not significant [$F(2, 330) = .903$, $p = .406$]. There was quite a bit of variation in reaction times within each episode type, and this may explain why the differences in reaction times between the three types of episodes were not significant.

Table 9 Reaction times in error detection on the stimulated correction task

	M (seconds)	*SD*	Minimum value	Maximum value
Error+recast	2.24956	0.66770	0.96560	4.65070
Error-recast	2.28148	0.70665	1.20420	5.40670
Correct	2.16076	0.70545	0.91940	4.90460
				p = ns

Note. ns = nonsignificant

Reaction times for error+recast and error-recast episodes were examined in relation to students' modifications of their errors, as shown in Table 10. The reaction times for error+recast and error-recast episodes were very similar when the student produced unsuccessful modification or no modification (2.24476 sec. and 2.24373 sec. respectively), but were faster on error+recast episodes when students modified their utterance in the same way as in the recast. These results suggest that the recasts may have led to faster reaction times for some of the episodes, namely those in which the students used the information from the recast to modify their utterance during the error correction task. However, a one-way ANOVA revealed that the differences in reaction times for error+recast/same as recast, error+recast/different than recast, and error-recast episodes were not statistically significant [$F(2, 68) = 1.141$, $p = .326$].

Table 10 Mean reaction times by nature of modification for error+recast and error-recast episodes

	Successful & partial		Unsuccessful & not modified
	Same as recast	Different than recast	
Error+recast	2.14991	2.51741	2.24476
Error-recast		2.40696	2.24373
			p = ns

Note. ns = nonsignificant

5. Discussion

The aim of the present study was to investigate whether incidental, extensive recasts provided by a teacher in a small group outside a classroom were beneficial to adult ESL learners. The results of the error detection task showed a trend for students to detect more errors in the error+recast episodes than in the error-recast episodes; this may point to a benefit from recasts in terms of students' ability to perceive errors in their own speech. However, the difference between error detection rates on error+recast and error-recast episodes failed to reach statistical significance on either the stimulated correction task or the written test. One possible explanation for this was that there seemed to be an overall bias for students to think that their speech contained errors, and this bias was found across all three episode types, including correct episodes. This bias may have been partially responsible for the high error detection rates on the error+recast and error-recast episodes. These high rates, all over 70%, may have been close to ceiling, thus muting any beneficial effect from the recasts and leading to the nonsignificant difference between the error+recast and error-recast episodes.

It was also found that on both the stimulated correction task and the written test, students were able to successfully and partially successfully modify

more of their errors from the error+recast episodes than from the error-recast episodes. While the difference in modification patterns was a trend in the stimulated correction task, it produced a statistically significant difference on the written test. These findings show that recasts seemed to benefit students' ability to correct errors in their own speech. In addition, there is evidence that the specific content of the recasts may have assisted students in the modifications of their errors. These results are especially encouraging given that the recasts that students received in the present study were spontaneous and extensive.

As for the students' reaction times, the results show that students responded slightly more quickly to error+recast episodes than to error-recasts, but this difference was not significant. The fact that students were able to respond more quickly indicates that students were possibly more confident (either consciously or unconsciously) in their answers to error+recast episodes than to error-recast episodes. While this trend did not reach significance, this was likely due at least in part to the great amount of variability in reaction times between students. There was also a trend for students to respond more quickly in the error detection task if they later went on to provide a successful or partially successful modification that was the same as the recast they had been given during the interaction. It is possible that the recast provided the day before was stored in a location that was "easily accessible," at least for a short period of time after it was provided.

The study also examined how students responded to correct episodes and how these responses compared to the error+recast and error-recast episodes. The data from the correct episodes showed that students perceived a considerable portion of their correct episodes as containing errors. On the stimulated correction task, students indicated that 47.7% of the correct episodes contained errors, while the rate for the written test was even higher, at 53.2%. These results indicate that students are often wrong in the assessment of the grammaticality of their own utterances, with a tendency to think that their utterances are less grammatical than they are. Results also showed that students often modified their grammatical utterances in correct episodes in such a way that they became ungrammatical.

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

beyond the error targeted by the recast. Specifically, the provision of recasts may also push students to question the grammaticality of their utterances that did not receive recasts. This possibility has yet to be examined by other researchers. In general, recast studies have only examined the effect of recasts on those individual erroneous utterances that received recasts (Loewen, 2005; Loewen & Philp, 2006; Nassaji, 2006, 2007) or on a target linguistic form (Ammar & Spada, 2006; Dilans, 2010; Ellis et al., 2006; Ishida, 2004; Lyster, 2004; Mackey & Philp, 1998; Saito, 2013).

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

The positive results of this study are particularly noteworthy in two respects. First, the recasts were beneficial even though they were provided in small-group (rather than dyadic) interaction. This indicates that recasts can be beneficial even when the teacher's attention is divided between several students. Secondly, the results of the present study show that recasts can be beneficial to students even when they are provided incidentally and extensively. Thus, while previous studies have demonstrated a benefit from intensive recasts (Ammar & Spada, 2006; Doughty & Varela, 1998; Ellis at al., 2006; Leeman, 2003; Lyster, 2004; Mackey, 2006; Mackey & Philp, 1998; McDonough & Mackey, 2006; Philp, 2003; Saito, 2013, among others), the findings of the present study demonstrate that recasts do not necessarily need to be provided intensively to be effective; even a single recast can be of benefit to students. These results confirm Loewen and Philp's (2006) and Nassaji's (2009) findings that spontaneous and extensive recasts can benefit students. As teachers often naturally provide incidental, extensive recasts in their classes, these results should be seen as especially encouraging for both researchers and teachers. Previous studies that have examined intensive recasts may have made teachers feel that recasts could only be beneficial if they were provided intensively. The present study shows that recasts can be effective when provided in response to a wide range of linguistic errors, even if some linguistic forms receive only one recast. As such, teachers should not be discouraged from incorporating spontaneous, extensive recasts into communicative-based oral interaction with their students.

6. Conclusions and implications

The present study has incorporated a number of innovative methodological features. First, in this study each student was tested on episodes that involved errors in their speech that did receive recasts, episodes that involved errors that did not receive any type of feedback, and episodes that did not involve any student errors. While a within-subject design may not be necessary for studies of recasts directed towards preselected target forms, such methodology might be very important in the study of spontaneous, extensive recasts. Previous posttest studies of spontaneous, extensive recasts (Loewen, 2005; Loewen & Philp, 2006) have not compared their test scores for errors that received recasts with anything. The present study, on the other hand, addresses the need for a control.

A second innovation of the present study is the use of *stimulated correction* as an instrument to measure learning. Stimulated correction was designed by the researchers and, to the best of our knowledge, is the first individualized posttest to make use of video clips in the assessment of learning following corrective feedback. Stimulated correction has several advantages over traditional written posttests, such as allowing students to see and hear the exact context in which the errors were made, and forcing students to make their judgments quickly, which may increase the chances that students are making use of their implicit knowledge to complete the task. This may be important given that it is often assumed that recasts are an implicit form of corrective feedback (Goo & Mackey, 2013).

A final significant contribution of the present study is that it was the first study in the area of recasts to examine student reaction time when completing posttests. While the reaction time results of the present study did not produce significant findings, it does not mean that the measurement of reaction time is not relevant to the study of corrective feedback. The measurement of reaction time has led to significant advancements in knowledge in fields such as psycholinguistics and will likely lead to important advancements in knowledge in SLA in the future.

Despite the contributions of the present study to the field of recast research, there are a number of limitations to the present study. First, the number of students who participated in the study was relatively small, and these small numbers may have been (at least partially) responsible for the lack of significant findings on some of the measures despite clear trends for a benefit from recasts on all of the measures.

Another limitation of this study is that it was conducted in a small-group environment (as opposed to a classroom environment). While this likely reflected a much more natural situation than that of dyadic studies, we must keep in mind that the results cannot necessarily be applied to a whole-class situation. This

being said, within L2 classrooms, teachers often divide students into small groups to complete tasks, projects, etc.

Another limitation of the study was the fact that all the students completed the stimulated correction task before the written test. The primary goal of this study was to examine students' responses on the new methodology, stimulated correction. Since the number of students in the present study was not large, it was decided not to further divide the students into two groups to receive the stimulated correction task and the written test in a cross-balanced manner. While a practice effect could have been in place on the written test, it should be remembered that the purpose of the test was not to see how accurate students were on error+recast episodes alone but to compare how accurate they were on error+recast and error-recast episodes. It was assumed that any practice effect would equally affect the error+recast, error-recast, and correct episodes. In addition, since the recasts were cut out of the video clips, the possibility that students would learn from hearing the recasts a second time was avoided.

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

Acknowledgments

This study is based on the first author's MA thesis (Hawkes, 2007) completed at the University of Victoria, Victoria, BC. We would like to thank the students who participated in this study and also the manuscript reviewers for their useful comments. This research was supported by a research grant from the Social Sciences and Humanities Research Council of Canada (SSHRC) to the first author.

References

Ammar, A., & Spada, N. (2006). One size fits all? Recasts, prompts, and L2 learning. *Studies in Second Language Acquisition, 28*, 543-574.

Crookes, G., & Gass, S. (Eds.). (1993). *Tasks and language learning.* Clevedon: Multilingual Matters.

Dilans, G. (2010). Corrective feedback and L2 vocabulary development: Prompts and recasts in the adult ESL classroom. *Canadian Modern Language Review, 66*, 787-816.

Doughty, C., & Pica, T. (1986). 'Information gap tasks': An aid to second language acquisition? *TESOL Quarterly, 20,* 305-325.

Doughty, C., & Varela, E. (1998). Communicative focus on form. In C. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 114-138). New York: Cambridge University Press.

Ellis, R., Basturkmen, H., & Loewen, S. (2001). Learner uptake in communicative ESL lessons. *Language Learning, 51*, 281-318.

Ellis, R., Loewen, S., & Erlam, R. (2006). Implicit and explicit corrective feedback and the acquisition of L2 grammar. *Studies in Second Language Acquisition, 28,* 339-368.

Ellis, R., & Sheen, Y. (2006). Reexamining the role of recasts in second language acquisition. *Studies in Second Language Acquisition, 28,* 575-600.

Farrar, M. (1990). Discourse and the acquisition of grammatical morphemes. *Journal of Child Language, 17,* 607-624.

Goo, J., & Mackey, A. (2013). The case against the case against recasts. *Studies in Second Language Acquisition, 35*, 127-165.

Han, Z. (2002). A study of the impact of recasts on tense consistency in L2 output. *TESOL Quarterly, 36,* 543-72.

Havranek, G. (1999). The effectiveness of corrective feedback: Preliminary results of an empirical study. *Acquisition et Interaction en Langue Étrangère, 2,* 189-206.

Hawkes, L. (2007). *Recasts revisited: The role of recasts in error detection and correction by adult ESL students*. (Unpublished MA thesis). University of Victoria, Victoria BC, Canada.

Ishida, M. (2004). Effects of recasts on the acquisition of the aspectual form *–te i-(ru)* by learners of Japanese as a foreign language. *Language Learning, 54*, 311-94.

Iwashita, N. (2003). Negative feedback and positive evidence in task-based interaction: Differential effects on L2 development. *Studies in Second Language Acquisition, 25,* 1-36.

Johnson, K. (1981). The four skills in communicative language teaching: Writing. In K. Johnson & K. Morrow (Eds.), *Communication in the classroom* (pp. 93-107). London: Longman.

Leeman, J. (2003). Recasts and second language development. *Studies in Second Language Acquisition, 25,* 37-63.

Li, S. (2010). The effectiveness of corrective feedback in SLA: A meta analysis. *Language Learning, 60*, 309-365.

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. (2005). Incidental focus on form and second language learning. *Studies in Second Language Acquisition, 27*, 361-386.

Loewen, S., & Philp, J. (2006). Recasts in the *adult* English L2 classroom: Characteristics, explicitness, and effectiveness. *The Modern Language Journal, 90,* 536-556.

Long, M. H. (1981). Input, interaction, and second language acquisition. In H. Winitz (Ed.), *Native language and foreign language acquisition* (pp. 259-278). New York: New York Academy of Sciences.

Long, M. H. (1983). Native speaker/non-native speaker conversation and the negotiation of comprehensible input. *Applied Linguistics, 4,* 126-141.

Long, M. H. (1996). The role of the linguistic environment in second language acquisition. In W. Ritchie & T. Bhatia (Eds.), *Handbook of research on second language acquisition* (pp. 413-468). New York: Academic Press.

Long, M. H. (2006). *Problems in SLA.* Mahwah, NJ: Lawrence Erlbaum.

Lyster, R. (1998). Negotiation of form, recasts, and explicit correction in relation to error types and learner repair in immersion classrooms. *Language Learning, 48,* 183-218.

Lyster, R. (2004). Differential effects of prompts and recasts in form-focused instruction. *Studies in Second Language Acquisition, 26,* 399-432.

Lyster, R., & Ranta, L. (1997). Corrective feedback and learner uptake: Negotiation of form in communicative classrooms. *Studies in Second Language Acquisition, 20,* 37-66.

Lyster, R. & Saito, K. (2010). Oral feedback in classroom SLA: A meta-analysis. *Studies in Second Language Acquisition, 32*, 265-302.

Lyster, R., Saito, K., & Sato, M. (2012). Oral corrective feedback in second language classrooms. *Language Teaching, 46*, 1-40.

Mackey, A. (2006). Feedback, noticing and instructed second language learning. *Applied Linguistics, 27*, 405-30.

Mackey, A., & Philp, J. (1998). Conversational interaction and second language development: Recasts, responses, and red herrings? *The Modern Language Journal, 82,* 338-356.

McDonough, K., & Mackey, A. (2006). Responses to recasts: Repetitions, primed production, and linguistic development. *Language Learning, 56*, 693-720.

Nabei, T., & Swain, M. (2002). Learner awareness of recasts in classroom interaction: A case study of an adult EFL student's second language learning. *Language Awareness, 11*(1), 43-63.

Nassaji, H. (2006, July). *Explicitness of interactional feedback and its role in the learning of the targeted form in student-teacher interaction.* Paper presented at the Pacific Second Language Research Forum (PacSLRF), Brisbane, Australia.

Nassaji, H. (2007). Elicitation and reformulation and their relationship with learner repair in dyadic interaction. *Language Learning, 57,* 511-548.

Nassaji, H. (2009). Effects of recasts and elicitations in dyadic interaction and the role of feedback explicitness. *Language Learning, 59*, 411-452.

Nassaji, H. (2015). *The interactional feedback dimension in instructed second language learning: Linking theory, research, and practice.* London: Bloomsbury.

Nassaji, H. & Hawkes, L. (2006, October). *Focus-on-form in L2 classrooms and the role of language proficiency.* Paper presented at the Second Language Research Forum (SLRF), University of Washington, Seattle, WA, USA.

Nicholas, H., Lightbown, P. M., & Spada, N. (2001). Recasts as feedback to language learners. *Language Learning, 51,* 719-758.

Panova, I., & Lyster, R. (2002). Patterns of corrective feedback and uptake in an adult ESL classroom. *TESOL Quarterly, 36*, 573-395.

Philp, J. (2003). Constraints on "noticing the gap": Nonnative speakers' noticing of recasts in NS-NNS interaction. *Studies in Second Language Acquisition, 25,* 99-126.

Sabbagh, S. (1998). *Learners' reactions to feedback in an adult ESL classroom* (Unpublished master's thesis). Pepperdine University, Malibu, CA, USA.

Saito, K. (2013). The acquisitioned value of recasts in instructed second language speech learning: Teaching the perception and production of English /ɹ/ to adult Japanese learners. *Language Learning, 63*, 499-529.

Saxton, M. (1997). The contrast theory of negative input. *Journal of Child Language, 24,* 139-161.

Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics, 11,* 129-158.

Schmidt, R. (Ed.). (1995). *Attention and awareness in foreign language learning.* Honolulu, Hawai'i: University of Hawaii Press.

Sheen, Y. (2006). Exploring the relationship between characteristics of recasts and learner uptake. *Language Teaching Research, 10,* 361-392.

Sheen, Y. (2011). *Corrective feedback, individual differences and second language learning*. New York: Springer.

Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. Gass & C. Madden (Eds.), *Input in second language acquisition* (pp. 235-253). Rowley, MA: Newbury.

APPENDIX

Transcription conventions

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