EYE TRACKING AS A MEASURE OF NOTICING:
A STUDY OF EXPLICIT RECASTS IN SCMC

Bryan Smith, Arizona State University

This study investigated whether eye-tracking technology could be employed as a measure of noticing of corrective feedback (in the form of explicit recasts) during NS-NNS task-based synchronous computer-mediated communication (SCMC). Pairs of university-level learners of English (*n* = 18) engaged in a short chat interaction task with a native speaker, who provided intensive and explicit corrective recasts. Participants' eye gaze record was compared to that from a stimulated recall. Noticing events (increased visual attention) were compiled and compared from each technique to determine whether these two techniques yielded similar data. Noticing events from each technique were also compared to results of immediate and delayed post tests of the targeted items. Results confirm the strength of both measures as methods for measuring what learners notice in the corrective feedback during SCMC. Further, the eye tracking and stimulated recall data also suggest that although learners engage in similar amounts of viewing activity across recasts targeting various linguistic categories, they are able to notice semantic and syntactic targets more easily than morphological targets. Results are discussed in terms of eye tracking as a potentially valuable tool in exploring the nature of noticing in instructed SLA and also in terms of argued benefits of CMC for language learning.

Keywords: Eye Tracking, Corrective Feedback, Recasts, Noticing, CMC

INTRODUCTION

The construct of noticing can be said to be key in all cognitively-oriented approaches to SLA. There have been several approaches to explaining this construct in the SLA literature (Gass, 1988; Robinson, 1995; Schmidt, 1990, 1995, 2001; Tomlin & Villa 1994; Truscott & Sharwood Smith, 2011), and noticing is widely viewed as a critical factor that mediates L2 input and interaction driven learning (Gass, 1997; Gass & Varonis, 1994; Long, 1996). Schmidt (1990) operationally defined noticing as the availability for verbal report. According to Schmidt, attention controls access to awareness and is responsible for noticing, which is necessary and sufficient for the conversion of input to intake. Schmidt (1995) suggests that input that is not attended to cannot be held in short-term memory and, therefore, is not available for further processing. Researchers espousing a cognitive interactionist approach (IA) to SLA have borrowed widely from Schmidt's notion of noticing. Gass (1997), for example, argues, "an initial step in grammar change is the learner's noticing (at some level) a mismatch between the input and his or her own organization of the target language" (p. 28). That is to say, linguistic input from an interlocutor must first be noticed before it may begin to be integrated into the learner's interlanguage (IL). Additionally, noticing this gap may represent a higher mental activity than the simple noticing of L2 exemplars, in that the identification of a gap between two forms involves a deeper linguistic analysis.

Measures of Noticing

In exploring the relationship between noticing and SLA, it is crucial that we have an adequate measure of what learners notice and when. Three methods have been employed most widely. One is the analysis of immediate uptake in the discourse (Braidi, 2002; Lyster & Ranta, 1997; Mackey & Philp, 1998; Tarone & Bigelow, 2007). The other two methods are types of verbal reports, including concurrent verbal reports such as think alouds (Alanen, 1995; Lai, Fei, & Roots, 2008; Leow, 1997, 2000; Rosa & O'Neill, 1999; Sachs & Suh, 2007) and retrospective verbal reports such as stimulated recall (Egi, 2007; Gass & Mackey, 2000; Mackey, 2006; Mackey, Gass, & McDonough, 2000; Mackey, Philp, Egi, Fujii, & Tatsumi, 2002).

Stimulated Recall

Stimulated recall is an introspective/retrospective means of exploring L2 learning processes which otherwise are not observable (Faerch & Kasper, 1987). In stimulated recall, prompts such as audio- or video-taped events are used to stimulate the participants' recall of their mental processes during the events. A well-executed stimulated recall approach to determining what learners notice (awareness at the level of noticing) in the input seems to be a strong choice for researchers in that it provides complementary data for understanding, for example, the effect of corrective feedback (Egi, 2007; Mackey, 2006; Mackey et al., 2002; Mackey et al., 2000; Egi, 2010).[1]

Recasts

The present study focuses on corrective feedback in the form of recasts. Therefore, a brief discussion of this important construct as it relates to the current study is in order. Brown and Bellugi (1964), first referred to recasts in the context of child L1 acquisition. And over the past fifteen years recasts have become one of the most widely studied types of corrective feedback in the second/foreign language (L2) classroom (Braidi, 2002; Doughty, 1994; Lyster & Ranta, 1997; Nabei & Swain, 2002; Oliver, 1995; Sheen, 2004). Though there are several competing definitions of recasts in the instructed SLA literature, recasts are essentially discourse moves (by an interlocutor) that, within the context of a communicative activity, correctly restate or rephrase (modify) all or part of a learner's utterance to be more target-like by changing one or more sentence components while still retaining its central meaning (Lyster & Ranta, 1997; Nicholas, Lightbown, & Spada, 2001; Sheen, 2006; Trofimovich, Ammar, & Gatbonton, 2007).

Recasts are generally theorized to be a powerful and potentially effective form of corrective feedback because they simultaneously provide both negative feedback and positive input (Leeman, 2003), arguably heightening the salience of targeted items (which may otherwise lack salience for learners) in a way that does not disrupt the flow of communication.

Though there are studies that show limited or no effects of recasts on SLA (Lyster, 2004; Lyster & Ranta, 1997; Ellis, Loewen, & Erlam, 2006; Loewen & Erlam, 2006; Sauro, 2009), a more compelling case can be made for their generally facilitative role in SLA (Goo & Mackey, in press), with their effectiveness being more pronounced in lab over classroom settings (see Ellis & Sheen, 2006; Li, 2010; Long, 2007; Lyster & Saito, 2010; McDonough & Mackey, 2006; Mackey & Philp, 1998; Mackey & Goo, 2007; Nicholas et al., 2001). Indeed, Nicholas et al. (2001) suggest that recasts appear to be most effective in contexts where it is clear to the learner that the recast is a reaction to the accuracy of the form, and not the content of the original utterance, something more likely in lab settings.[2]

Problematizing Recasts

Comparing the research on recasts is problematic not least of all because recasts are a dynamic and multifaceted, rather than a one-dimensional construct. For example, Sheen (2004) points out that recasts are a bit more complex than most definitions suggest. Indeed, there are many different types of recasts, including corrective, non-corrective, intensive, explicit, implicit, full, partial, simple, complex, adjacent, and non-adjacent, among others. That is to say, recasts vary in their form and presentation thus making general conclusions about their efficacy challenging.

Factors Affecting the Noticing of Recasts [3]

Current research suggests that many factors affect the noticing of recasts. At a very basic level several researchers have argued that there is a great degree of ambiguity in teacher-provided recasts and therefore, recasts may be less effective than other types of feedback moves, such as prompts (Chaudron, 1977; Fanselow, 1977; Lyster, 1998; Lyster & Ranta, 1997; Nabei & Swain, 2002). One of the main sources of this ambiguity—especially in the communicative and meaning-based classroom—is that students are expected to sort out whether the teacher's intentions are focusing on form or meaning. In

addition to the argued general ambiguity of recasts, other factors that may affect the noticing of recasts include the nature of the recast itself (Mackey et al., 2002; Long, Inagaki, & Ortega, 1998; Trofimovich et al. 2007), learner-internal variables such as proficiency level and/or learner readiness (Ammar & Spada, 2006; Mackey & Philp 1998; Philp, 2003), working memory capacity (Mackey et al., 2002; Sagarra, 2007; Tarone & Bigelow, 2007), and learner anxiety (Sheen, 2008). Factors related to the pedagogical and social context such as learner familiarity with the type of feedback provided and setting are also important to consider.

When considering the relationship between the nature of the recast itself and noticing the following aspects are apparent: the form targeted by the recast (the linguistic category) and the structural composition and linguistic context of the recast. In terms of the former, there is evidence that learners first focus on (and arguably notice) the semantics of interactional exchanges and only later focus on form (Mackey et al., 2000; Tarone & Bigelow, 2007) despite the fact that teachers seem more inclined to provide recasts more often to grammatical than to lexical errors (Lyster, 1998; Mackey et al., 2000; Nabei & Swain, 2002). This is likely the explanation for the higher effectiveness of the lexical recasts over grammatical recasts (Mackey et al., 2000; Sheen, 2006; Trofimovich et al., 2007; Williams, 1999).[4]

In terms of the structural composition and linguistic context of recasts, several variables are apparent, which may affect their effectiveness. For example, Ellis et al. (2006) suggest that recasts may lie on an explicit-implicit continuum. Recasts may also be declarative or interrogative (Kim & Han, 2007; Loewen & Philp, 2006). Finally, the length, number of changes, and number of recasts on the same target seem to be important factors related to effectiveness (Loewen & Philp, 2006; Philp, 2003; Tarone & Bigelow, 2007).

Setting and pedagogical context are extremely important in recast research (Ellis, Basturkmen, & Loewen, 2001; Morris & Tarone, 2003; Nabei & Swain, 2002; Nicholas et al., 2001; Oliver & Mackey 2003; Sheen, 2004, 2006). Factors such as learners' prior familiarity with the input (Sheen, 2004), task type and task content (Nicholas et al., 2001; Révész, 2009), and learner perception (Morris & Tarone, 2003) have been argued to be key. Morris & Tarone (2003), for example, suggest that recasts may provide greater opportunities for language acquisition in contexts where both teachers and students are oriented to attending to linguistic form as opposed to focusing exclusively on meaning.

While it is certainly true that many learner internal and external variables will combine to affect the likelihood that recasts will be noticed by learners and, therefore, be potentially effective, methodological issues surrounding the capturing, coding, and evaluation of recasts must be accounted for as well. Indeed, Trofimovich et al. (2007) suggest that any lack of consensus regarding the relationship between recasts, noticing, and SLA is largely due to the different methodologies employed by researchers to measure noticing, as well as the use of various measures of working memory, proficiency level, etcetera.

Written Recasts

Seedhouse (1997) suggests that teachers often prefer recasts to other forms of corrective feedback because they are relatively non-threatening, mitigated, unobtrusive, and implicit in nature. However, this assumption is by no means absolute. The assumption that recasts are necessarily implicit in nature, for example, is perhaps too simplistic for an instructed SLA context (Ellis & Sheen, 2006; Sheen, 2006) since they may surely lie at various points on a continuum of linguistic implicitness-explicitness (Ellis et al., 2006). For example, in an oral-aural mode accurately positioning a recast on this continuum may be affected by things like repetition and stress (Doughty & Varela, 1998). One might also argue that written corrective recasts in a synchronous computer-mediated communication (SCMC) environment are necessarily more explicit than those found it its oral counterpart (yet still less explicit than, say, direct error correction), all other things being equal (see also Dasse-Askildson, 2008). This is due to the written nature of the medium itself as well as the message permanence afforded. Similarly, the argued ambiguity (see discussion above) of oral recasts seems less convincing in the written mode. That is to say, written

recasts may be more forgiving of the learner's need to simultaneously distinguish features of the interlocutor's responses that sustain the interaction from those that provide a means of focusing on the language used in the interaction (see Nicholas et al., 2001). It also seems unlikely that learners will interpret a written recast as simple repetition of their previous (well-formed) utterance.

Recasts in Text-Based SCMC

The research on recasts in a text-based CMC environment is quite limited. Loewen and Erlam's (2006) investigated the relative effectiveness of recasts and metalinguistic prompts administered during small group text-chat interaction. They found no significant advantage for either feedback type over the control condition and no significant advantage for one corrective feedback type over the other. Sachs and Suh (2007) compared the efficacy of textually enhanced and unenhanced CMC recasts in the development of a targeted form (back shifting of verbs from simple past to present perfect in reported speech). They found that textual enhancement was related to reported awareness and that higher levels of reported awareness showed stronger correlations with posttest performance. However, there was no direct significant relationship between enhancement and posttest performance. Sauro (2009) used full recasts in comparison with metalinguistic feedback. Neither feedback type was significantly more effective than the other in either the immediate term or over time. Lai et al. (2008), using think aloud and stimulated recall to measure noticing, suggest that the contingency of CMC recasts affect whether or not they are noticed. This explanation is related to the notion of enhanced salience (Leeman, 2003) with more contingent recasts (perhaps) being more salient than non-contingent recasts. In Lai et al., 53% of the contingent recasts were noticed as opposed to 35% of the non-contingent recasts.[5] In a study employing a similar methodology to that reported in the current one, Smith (2010) showed that learners noticed about 60% of the intensive recasts they received. Lexical recasts were found to be much easier than grammatical recasts for students to notice, retain, produce more accurately on a written posttest, and use more productively in subsequent chat interaction. Similarly, Smith and Renaud (in press) found that learners were able to pay careful attention to both lexical and grammatical targets in recasts. Learners focused on close to 75% of teacher recasts with between 20% and 33% of these resulting in posttest gains.

There are many potential benefits of text-based CMC for SLA. From an interactionist approach, which informs the present study, these benefits include increased student participation, an increased quantity and heightened quality of learner output, an enhanced attention to linguistic form, and an increased willingness to take risks with their second language (Smith, 2004; Pellettieri, 1999). The slower speed of typing as well as a software- and network-induced lag time between turns, coupled with the heightened salience of input (including corrective feedback) and output afforded by the permanence of the written message on the screen means that interlocutors have more time to both process incoming messages (Pellettieri, 1999) and produce and monitor their output (Smith, 2008; Sauro & Smith, 2010). Essentially, then, from a cognitive interactionist perspective, the potential advantages afforded SLA by text-based SCMC comes down to the construct of noticing—especially the noticing of corrective feedback, as well as non-target like input and output (Smith, 2004; Izumi, 2003; Salaberry, 2000). More specifically, text-based SCMC may be said to afford more online planning time, which is argued to contribute to L2 performance by freeing up attentional resources (Yuan & Ellis, 2003). Thus learners are able to more closely attend to linguistic form while formulating messages and also engage in more focused pre- and post-production monitoring (Yuan & Ellis, 2003). Sauro and Smith (2010), using screen capture technology in an SCMC study, convincingly show that learners do appear to use the increased online planning time afforded by chat to engage in careful production that results in more complex language. Though the actual online planning itself remains an unobservable activity (Skehan & Foster, 2005), their findings support previous (face-to-face) studies (Ellis & Yuan, 2004; Yuan & Ellis, 2003), which also found a positive effect for increased online planning time. Recasts, then, in an SCMC environment might be expected to be especially salient to learners, though the research to date seems mixed (Lai et al. 2008; Sauro, 2009).

Eye Tracking

Before moving on to a discussion of the role eye tracking played in the current study, a brief examination of eye tracking technology in general is in order. Eye tracking technology has been employed as a tool in psychological reading research for over 100 years. It consists of a suite of techniques whereby an individual's eye movements are measured so that the researcher knows both where a person is looking at any given time and the sequence in which their eyes are shifting from one location to another (Poole & Ball, 2006).

Eye movements during reading—such as gaze duration, saccade length, occurrence of regressions, and search time—can be used to infer moment-by-moment cognitive processing of a text by the reader (Just & Carpenter, 1980) without significantly altering the normal characteristics of either the task or the presentation of the stimuli (Dussias, 2010). These eye movements are considered empirical correlates of processing complexity, which allow us to make inferences about perceptual and cognitive processes. For example, the analyses of eye-movement patterns have been used to explore the cognitive processes of pronoun resolution and co-reference, word frequency, lexical ambiguity, syntactic ambiguity, and discourse factors (for a review, see Rayner, 1998).

The two most widely used measures of eye movements are eye fixations and saccades. Eye fixations are those moments when the eyes are relatively stationary and reflect when information is being encoded. It is eye fixations that allow readers to extract important and useful information about the text (Dussias, 2010). Saccades are the fast movements of the eyes between fixations. No encoding takes place during saccades, so they are not helpful in interpreting a target's complexity or salience.[6] In general, eye fixations during (L1) reading in English last approximately 200–250 milliseconds (Rayner, 2009), though there is considerable within- and between-reader variability. For example, a fixation may range from just under 100ms to more than 500ms. This variability in fixation duration is believed to vary as a function of the relative cognitive difficulty in comprehending a text (Rayner, 1998). However, L1 readers do not fixate on every word in a text, but rather they fixate on about two-thirds of the total words (Just & Carpenter, 1980).

Context is extremely important in interpreting the meaning of fixations, which cannot be done in a vacuum. For example, when browsing a Web page (an encoding task), higher fixation frequency on a specific area can mean greater interest in the target, such as when one views a picture in an advertisement. Higher frequency can also be a sign that the target is complex in some way and more difficult to encode (Jacob & Karn, 2003; Just & Carpenter, 1976). Poole, Ball, and Phillips (2004) suggest that more fixations on a particular area indicate that it is more noticeable, or more important, to the viewer than other areas (reported in Poole & Ball, 2006). However, in search tasks, a higher number of single fixations, or clusters of fixations, are often an indicator of greater uncertainty in recognizing a target item (Jacob & Karn, 2003). The duration of a fixation is also linked to the processing-time applied to the object being fixated with a longer fixation duration indicating either difficulty in extracting information, or that the object is more engaging in some way (Just & Carpenter, 1976). The so-called eye-mind assumption holds that the reader's eyes remain fixated on a word as long as the word is being processed.

Up to now, eye-movement records in L2 research have been used to investigate two major areas: (a) the way in which L2 speakers recognize words when they are spoken in each language; and (b) the question of whether monolingual and L2 speakers process various syntactic sub-processes similarly during sentence comprehension tasks (see Dussias, 2010, for an overview). Only a handful of published studies have so far employed eye tracking to explicitly examine L2 noticing (Godfroid, Housen, & Boers, 2010, Kuhn, 2012; O'Rourke, 2008, in press; Smith, 2010; Smith & Renaud, in press). Each of these studies suggests that eye tracking is suitable to use as an instrument for measuring the noticing of written text.

Justification of this Study

One basic assumption of this study is that linguistic input from an interlocutor must first be noticed before it may begin to be integrated into the learner's interlanguage. A second assumption is that the written nature and slower pace of the SCMC medium is likely to facilitate noticing, in this case the noticing of corrective feedback in the form of explicit recasts. Finally, eye gaze data is assumed to provide a relative indication of processing time applied to the object being fixated in relation to other relevant text. It is also assumed that the eye tracking technique is unlikely to interfere with the participants' cognitive processing (Godfroid et al., 2010).

The present study has three purposes. First, I wanted to capitalize on the affordances of computer technology in order to examine what learners attend to in explicit recasts. To do this, one of the most common measures of noticing was employed (stimulated recall) as well as a second measure based on heat maps created while tracking participants' eye gaze during task completion. Further, since eye tracking is an (alternative) online measure, it may be more robust than retrospective techniques such as stimulated recall. The recent use of eye tracking technology in explorations of learner noticing during SCMC activities seems to confirm findings emerging from the face-to-face literature about the relationship between the form targeted by the recast and subsequent noticing and retention, with lexical items being advantaged over grammar (Smith, 2010; Smith & Renaud, in press). In the present study, scores from tailored posttests were used as a dependent variable with evidence of noticing indicated by the two measures above (eye gaze and stimulated recall) serving as independent variables.

The second purpose of this study was to determine the extent to which stimulated recall and eye gaze correlate with one another. This is in light of suggestions made earlier that it is the methodology used to measure noticing that contributes to the lack of clear agreement on the effectiveness of recasts argued by some in the literature. Third, I wanted to further explore the extent to which the variable linguistic category mediates the noticing of recasts in a written CMC environment.

In order to further explore the issues above, the following research questions were posed:

1. What is the relationship between noticing of the targeted recast items (as measured by noticing events in the heat map and stimulated recall records) and posttest scores for these same target items?

2. What is the relationship between heat map and stimulated recall to one another when used as measures of noticing?

3. What is the relationship between the linguistic category of the recast target and noticing?

METHODOLOGY

Participants

Pre-Treatment: Pre-Test

Potential participants were recruited by a general call for volunteers circulated among the non-native speaker sections of freshman composition as well as among the university's intensive English program and international student groups at a large university in the United States. Volunteers were paid a small amount for their participation in the study. Participants' ($n = 18$) self-reported L1s were as follows: Chinese (10), Cantonese (2), Korean (2), Turkish (2), Mandarin (1), and Thai (1).

Instruments and Procedures

Pre-Treatment: Pre-Test

A link to an online test of general English proficiency was sent to each participant prior to the treatment (http://www.transparent.com/learn-english/proficiency-test.html). Participants completed 40 four-item multiple choice questions about English grammar and vocabulary. Raw scores are automatically

calculated by the website upon completing the test and scores are immediately available to the user. These proficiency test scores, which ranged from 63–98 (out of 100) were then recorded and their overall score was used in testing for normal distribution across this variable, which was sufficiently met ($n = 17$,[7] $M = 86$, $SD = 11$).[8] Figure 1 shows the order of each of the procedures discussed in this section.

Proficiency Test	Self-led introduction to think alouds and recasts
Day of treatment	Assessment of recast knowledge Calibration of eye tracker Treatment task Immediate post test Stimulated recall
Post treatment	Delayed post test

Figure 1. Procedures.[9]

Recast Training

I discussed earlier the multiple potentially confounding variables in interpreting recast research. Among these variables is the notion of ambiguity, which is inextricably intertwined with pedagogical and social context. One major design aspect of the current study was to remove this variable to the extent possible. To this end, participants were explicitly trained on the nature and intent of recasts, which in this case were designed to orient participants to form rather than meaning.[10]

A few days prior to their treatment session, a sheet detailing the concept of recasts in language teaching was sent to each participant. This sheet also included examples of several types of recasts, lexical, morpho-syntactic, semantic, etcetera. Participants were asked to review this sheet and familiarize themselves with the concept of recasts as well as with the examples provided. A separate recast evaluation sheet was also sent to each participant. This sheet consisted of six items that reflected a short fictitious teacher-student classroom exchange. Participants were asked to determine whether each item reflected a recast or not.

The conscious choice to explicitly introduce learners to the form and function of recasts was based on research that shows learners' orientation toward and familiarity with specific types of corrective feedback is an important variable in their ability to utilize the feedback (Morris & Tarone, 2003; Philp, 1999; Sheen, 2004), and also on the suggestion that recasts appear to be most effective when it is clear to learners that the recast is a reaction to the accuracy of the form and not the content of their original utterance (Nicholas et al., 2001). It is my view that in an instructed L2 setting, learners do indeed gradually (or even quickly) pick up on the instructor's intent when providing corrective feedback of various types, as well as when they engage in questioning, praising, silence, etcetera. As such, the notion that a particular form of corrective feedback is implicit, explicit, or somewhere in between is largely one stemming from the instructor's point of view rather than that of the learner. The trade-off in making this methodological decision, of course, is that the relevance of any findings to studies of purely implicit recasts or those conducted in markedly different pedagogical (and social) contexts may be limited. Nevertheless, I argue that the approach taken here in no way compromises the argued benefit of recasts, which essentially comes down to teachers heightening the salience of a targeted form by providing both positive and negative evidence in a way that is palatable to the communicative classroom.

Day of Treatment

Participants handed in their pretest sheets as well as their recast evaluation sheet at their scheduled session. If participants scored less than 100% correct on their recast sheet, the researcher reviewed those

incorrect items together with them and explained those items that were incorrect. Learners were allowed to ask questions in order to clarify the nature and purpose of recasts in their minds.

Next, the researcher verbally demonstrated two additional recasts for students and discussed with each participant why each of these was a recast. The participants were then asked to think for a moment and come up with a short dialog where participant A was the student and B was the teacher. This dialog was to contain one error of some sort and one recast by the teacher. Participants were then asked to identify the error and the recast. This part of the procedures concluded with a short discussion of whether recasts are common in language learning classrooms in the participant's home country as well as their experience with recasts in ESL classrooms in the United States. In this way, it was reasonable to assume that participants had a good grasp of the nature and purpose of recasts, therefore, removing a potentially confounding variable from the research design. That is to say, every effort was made to ensure that participants were aware of the existence, intent, format, and function of recasts as a form of corrective feedback.

Calibrating the Eye Tracker

A MangoldVision portable eye tracker from EyeTech Digital Systems (model TM3) was used in this study along with the MangoldVision Gaze Data Analysis Software Suite. This remote eye tracker fastens to the bottom edge of any computer screen.[11] An infrared camera follows one's pupil movements and can record one's eye gaze in real time. It also allows for a normal range of head movements (25cm x 16 cm x 19 cm) and will find one's pupils again even if one is temporarily out of range. The eye tracker must be individually calibrated to each learner. Learners follow a red ball displayed on the screen with their eyes for about 30 seconds until the software provides a score reflecting the calibration for each eye. For this study, the acceptability level was set at 75%. In order to provide an additional source of calibration, learners were instructed to slowly focus on a test line that consisted of three specific letters (spaced apart from one another), which appeared as the first message from the researcher at the beginning of the chat session. They were also asked to read these letters aloud in order to provide additional verification regarding each learner's eye gaze patterns. That is, it was possible to match the eye gaze record with the audio record of where participants were focusing their attention.

In addition to the eye tracking data captured, the Mangold software includes a video (and audio) screen capture capability, which was used to play back the chat interaction (with the eye gaze tracking ball disabled) in real time during the stimulated recall sessions. Camtasia 6.0 was used to capture the audio and video playback of the stimulated recall sessions. The text-based chat software used in this study was PSI, a jabber-based compose and post software. The spell check (check as you go) function was disabled.

The Task

Learners were asked to view a short clay animation video clip (Plonsters) of about 2½ minutes in duration. The Plonsters characters use no real language in the clip, though they do communicate with one another using an indecipherable gibberish. These characters also use a wide range of facial expressions and extra- and para-linguistic vocalizations.[12]

Immediately after the video ended, participants were asked to chat with their interlocutor online. The task goal was to retell the story as accurately as possible to allow the interlocutor to put a sequence of still images taken from the video into the proper order (sequential ordering task). Learners were told that their interlocutor may ask questions as well as provide comments and corrective feedback including recasts.

In additional to providing continuers (such as "OK"), comments, and follow-up questions on the content of the learner's previous message, the researcher (interlocutor) provided full recasts when it seemed natural to do so.[13] These recasts were provided at the earliest possible point in the chat interaction after the researcher noted a problematic utterance, thus resulting in recasts that were primarily contingent in nature (Lai et al., 2008). Further, since the interlocutor stressed the naturalness of the recasts, many of

these recasts targeted more than one linguistic item in the participant's previous message. During the task, all text-based chat interaction was logged automatically to a text file. Once the task was declared finished by the learner, the chat file was saved and the MangoldVision Gaze program stopped. A maximum of 12 minutes were allowed for the chat task.

Immediate Posttest (IPT)

Immediately following the chat session, learners were asked to write (in MS Word) a narrative retelling of the video they had just seen. They did so using a laptop PC set up in the same room. Spell and grammar checking were disabled on the computer. Participants were given 12 minutes to complete this test. No outside resources were allowed.

Stimulated recall

Upon completing the immediate posttest, participants engaged in a stimulated recall with the researcher. The chat interaction was replayed using the MangoldVision Gaze Data Analysis Software, without sound and with the eye tracking ball disabled. As part of the stimulated recall procedure, the video playback was paused after each incoming message from the researcher. Camtasia Studio 6 was used to record the stimulated recall session, including the screen capture of the chat interaction and the audio produced during the stimulated recall questions and answers. At each message the researcher asked "What were you thinking when I sent you this message?" This was often followed up by the question "At the time, did you notice anything in particular about my message?" When learners indicated that they had in fact noticed some difference between their preceding message and the recast (where applicable), they were asked to specifically identify this difference verbally. The researcher also asked specifically if they noticed that difference "just now" or "during the chat."

Delayed Posttest (DPT)

The tailored delayed posttest consisted of a participant-specific chat transcript adapted to isolate the participant's output. For the DPT, the chat transcripts could be reviewed by the researcher and each targeted (recast) item could be presented to learners, something not possible in the immediate posttest. Asking participants to complete a written story retelling similar to that in the IPT was deemed unrealistic given the short exposure to the story (video) and the time lag between posttests. All messages from the chat session that were sent by the researcher/interlocutor were removed. That is to say, the DPT was a transcript of the learner's half of the chat interaction. Each substantive line of learner output was numbered. Complete sentences from the chat transcript that were spread across multiple lines during the chat interaction were combined into one cohesive line for the purposes of the delayed posttest. The number of entries on the delayed posttest varied for each participant based on the length of their chat transcript. Since all lines of learner transcript text were included in the delayed posttest, those lines that were not the target of a recast during the chat interaction (well-formed utterances) served as distractors. The delayed posttest asked learners to determine whether each sentence was correct as written or if there were one or more errors in the sentence. If participants felt the sentence was correct as written, then they were instructed to simply write "OK" under the sentence. If they felt that there were one or more errors, they were instructed to write "No" and then rewrite the sentence in its entirety repairing the error(s). Thus, the DPT accounted for the target items of every recast offered by the researcher. The DPTs were e-mailed to each student one week after the student's treatment session. They were asked to complete the posttest within one day and e-mail it back to the researcher.[14] Once the DPT file was opened participants were given one minute per item to complete the DPT. That is, if a learner's DPT consisted of seventeen items (say, ten target items and seven distractors), this learner would have seventeen minutes to complete the DPT.[15]

Data Coding and Analysis

Recasts

Recasts from the native speaker interlocutor were always sentence-length and often involved more than one linguistic change. Though this complicates interpretation of the results in some ways, this approach was deemed more ecologically valid and is not without precedent in the literature (Sachs & Suh, 2007). Indeed, using Sheen's categories above, the recasts found in the present study would generally be characterized as corrective, relatively intensive, explicit (not least of all because of the nature of the medium), full, both simple and complex, and typically adjacent. There were 114 recasts total in the data set reflecting 156 specific linguistic targets. This translates into about 1.37 targets per recast. The number of recast targets per participant ranged from 2 to 16 with a mean number of about 8. All targeted items were then coded for linguistic category. Though there are many approaches to coding corrective feedback, the approach and coding scheme suggested by Liu (2008) was adopted for its balance of practicality and precision (see Figure 2). This resulted in three categories for linguistic feedback: morphological, semantic, and syntactic. A second trained rater used the same coding scheme and coded one-half of the transcripts. An inter-rater reliability analysis using the Kappa statistic was performed to determine consistency among raters. The inter-rater reliability for coding the Linguistic Category of each recast was found to be Kappa = 0.753 (p < 0.001), 95% CI (0.624, 0.882). This was deemed sufficient to proceed with the coding decisions reached by the main researcher.

Morphological errors:
> All errors in verb tense or form; plural or possessive ending incorrect, omitted, or unnecessary; subject-verb agreement errors; article or other determiner incorrect, omitted, or unnecessary.

Semantic errors:
> Errors in word choice, including preposition and pronoun errors; omitted words or phrases, unnecessary words or phrases. Spelling errors included only if the (apparent) misspelling resulted in an actual English word.

Syntactic errors:
> Errors in sentence/clause boundaries (run-ons, fragments, comma splices), word order, other ungrammatical sentence constructions.

Figure 2. Procedures for marking errors. From Liu, 2008.

Measures of Noticing

There were up to two different measures of noticing recorded for each recast item for every participant. I borrow the term *noticing event* from Godfroid et al. (2010) since both eye gaze and stimulated recall data are indirect indicators of noticing and tell us little about the nature of cognitive processing that ensues from such an event. Noticing events are essentially evidence of increased visual attention to the target, which suggests that the reader was cognitively engaging with the target item. In terms of the eye gaze data, this increased visual attention is marked by the overall dwell time, which consists of the sum of all fixations on a word, including first run dwell time and all regressions.

In contrast, noticing events during stimulated recall are defined here as cases where learners indicated that a specific recast was targeting an earlier non-target-like item and they were able to verbally identify the noted difference. I signify a noticing event with the symbol "+" and the absence of a noticing event with the symbol "-". For the statistical analysis discussed later, those items reflecting a noticing event were coded as 1 whereas those items where no noticing event is evident were coded as 0. Further, since there will be many references to noticing events across both measures of noticing employed in this study, I will abbreviate these as follows:

Heat map   HM+   (noticing event present)
                         HM-   (no noticing event present)
Stimulated recall   SR+   (noticing event present)
                         SR-   (no noticing event present)

Heat maps

The Mangold software has the capability of creating heat maps for subjects, which take into account the number and duration of eye fixations on a particular part of the computer screen relative to the length of the eye gaze recording. With this particular software, clear areas of the heat map reflect no eye fixations. The relative number of eye gaze fixations is reflected by the main colors on the color spectrum beginning with purple (the fewest number of eye fixations) to blue, green, yellow, orange, and red (the highest number of eye fixations). After comparing the actual eye tracking playback (where the eye gaze tracking ball is visible in real time) to the test line (see explanation above), it was determined that using the color blue (or hotter) on the heat map was sufficient as the threshold for indicating an increased amount of attention to that specific part of the computer screen. In order to code a specific recast item as HM+, the salient part of the recast text had to be positioned under some amount of blue (or hotter) shaded area (see Figure 3 below). This liberal threshold allows for eye gaze calibration fluctuations across learners since it is not at all clear how many eye gaze fixations or their duration would constitute a sufficient number and length for noticing for any particular learner.[16] That is to say, in helping determine what learners notice in the input, the duration of an eye gaze fixation is rather arbitrary. Heat maps in this case provide a relative duration of eye gaze, which as discussed below begins when the relevant recast appears on the learner's screen and ends when the screen shifts position, for example when another line of text appears.

Each heat map is based on the time immediately following the appearance of the recast on the learner's computer screen (post pane) and runs until just before there is a shift in the post pane, such as when another incoming (or outgoing) message is posted (or in the rare case, scrolling up). Since this part of the research was not concerned with what learners were composing, the movement in the compose window did not affect the salient aspects of the heat map record. The exception to this was in the case of lengthy messages, whereby each pressing of the return (enter) key by the writer would cause a shift in the post pane. In this case, the heat map clip would stop immediately before this shift in the post pane occurred.
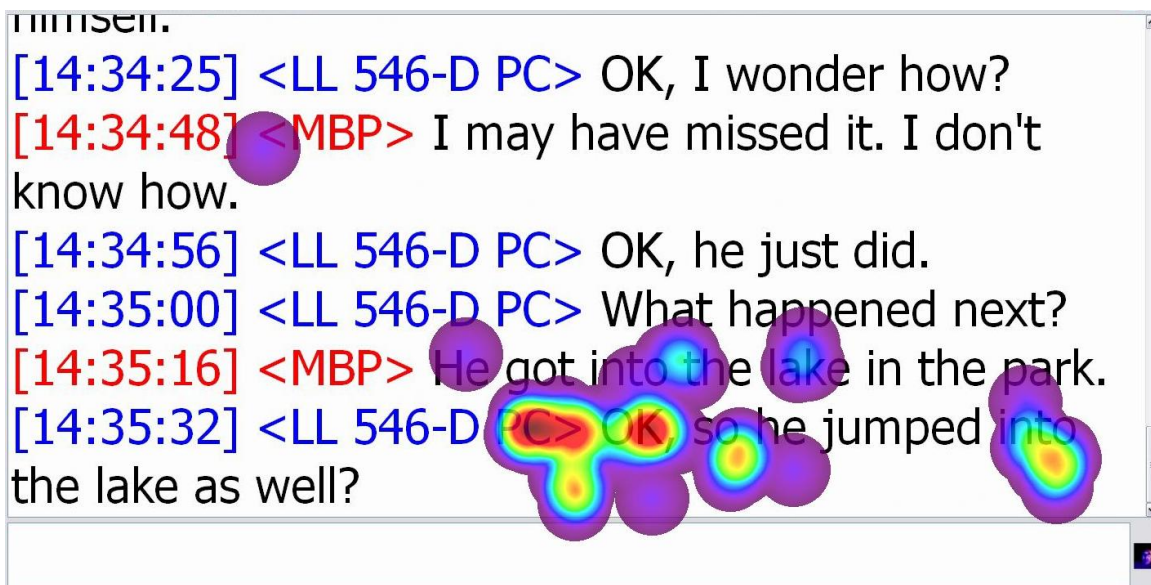

Figure 3. Heat map reflecting no noticing event.

In many cases it was necessary to create several heat maps for a single recast item to accommodate these shifts in the screen. Figure 3 (above) reflects a single item recast. The recast in this case was targeting the verb "got," recasting it as "jumped." Based on the heat map data, this item was coded as "0" (not noticed) for the variable heat map since there was no blue shaded area on any part of the word *jumped*.

In contrast, Figure 4 (below) shows a heat map for the single recast item targeting the form of "to throw" used by the learner. Since there is blue shaded area touching part of the word "threw," this recast item was coded as "1" (reflecting a positive noticing event) for the variable heat map.
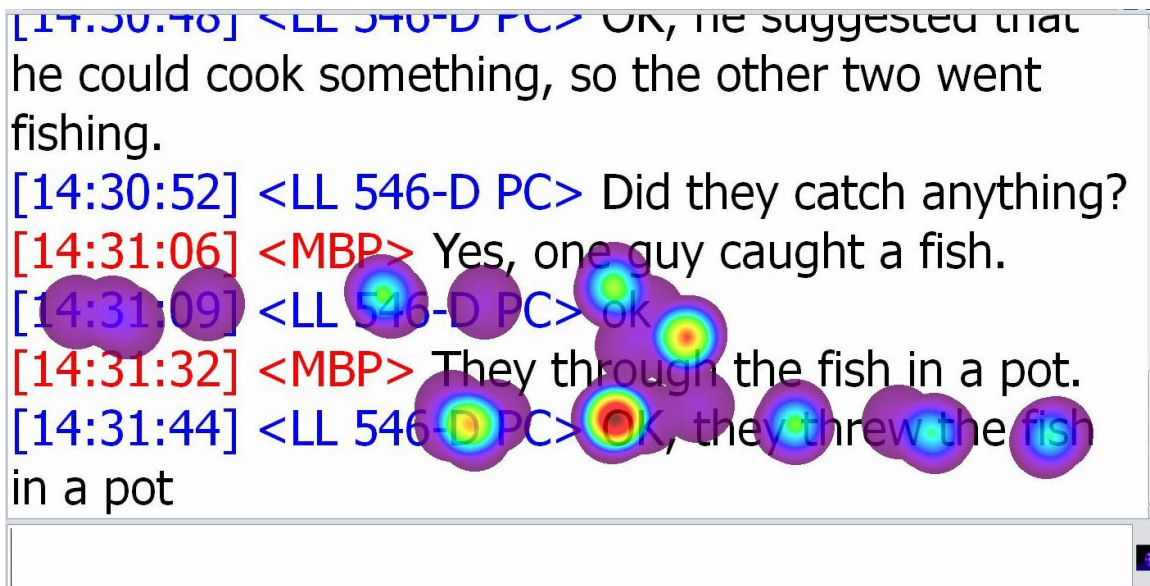

Figure 4. Heat map reflecting a positive noticing event.

Stimulated Recall

Like the heat map measure, stimulated recall was coded as either 0 or 1. In cases where learners indicated that a recast was targeting an earlier non-target-like item and were able to specifically identify verbally the noted difference, that item was coded as 1 (SR+). Again, such items were only coded as SR+ if learners specifically indicated (upon prompting) that they had noticed a specific difference between their previous output and the interlocutor's recast "during the chat." Those items where learners indicated that they just noticed the difference during the stimulated recall were coded as 0 (SR-). As will be discussed below, this presents some difficulty in analyzing the delayed posttest data for stimulated recall since stimulated recall essentially becomes part of the treatment (see also Swain, 2006).

Immediate Posttest (IPT)

Each IPT was reviewed for productive use of the targeted recast items from the chat interaction. Since recasts were only provided for non-target-like utterances, each targeted item was considered unknown (or at least not mastered) prior to the treatment. It is important to note that in cases where learners initially used an item in a target-like way, but then later in the chat used this same item in a non-target-like way (resulting in a recast from the interlocutor), this item was excluded from the data since it may reflect some degree of competence with that item. That is to say, only those items that showed only non-target-like use upon their first attempt by learners during the chat interaction are included in the data below. Further, if there was only one attempted use at an item and this use was also non-target-like and during the stimulated recall the learner revealed that they had known the correct form of the targeted item (performance error), then this item was also removed from the data.

Since the IPT essentially asked learners to repeat the treatment task, it was expected that a large percentage of the structures and items used during the chat interaction would be repeated during the IPT. Only those items that were recast during the chat were candidates for consideration. In cases where a single target item was attempted several times in the IPT, only those cases where all attempts were target-like or non-target like were considered and issued a score of 1.0 and 0.0 respectively. In cases where the IPT showed a mix of target-like and non-target-like use of a recast item, this item was removed from the IPT data. Not all targeted items were attempted by learners during the IPT. In the very few instances where learners made grammatically correct changes to the targeted item(s) that differed from the form suggested in the interlocutor's recast, these cases were omitted from the data. These coding decisions were made in view of the types of statistical analysis chosen, (which require binomial data) and are considered a limitation of the study.

Delayed posttest (DPT)

The DPT score for each targeted item was calculated in the following way: If learners failed to correct a sentence that was recast in the treatment phase, then this item was scored 0 (incorrect). Likewise, if they made non-target-like changes to a recast item, then this was also scored 0 (incorrect). Changes (correct or incorrect) made to the distractor items were ignored. Only those recast sentences that were rewritten to follow the form suggested by the interlocutor in the recast were scored as 1 (correct). In the very few instances where learners made grammatically correct changes to target items in a recast sentence that did not mirror that offered by the interlocutor during the chat session, these cases were omitted from the data.

RESULTS

The data discussed below are made up of 156 total target items across 18 participants. The actual number of recasts (114) was smaller than the 156 total target items since many of the recasts contained multiple targeted items. Though the actual number of errors addressed in any given recast is certainly of interest, this study focused on full recasts wherever possible and, therefore, contained both single item and multiple item recasts. The difficulty in providing only partial recasts is that learners may take the interlocutor's omission of corrective feedback to be a tacit confirmation of the accuracy of their non-target-like use.

Table 1 below shows the noticing events across the two measures explored. This table shows that where stimulated recall occurred for an item (where there was a valid record of stimulated recall for that item), learners indicated noticing (SR+) that targeted item during the chat interaction in 74 of the 137 cases, or 54% of the time. No noticing SR- was reported in 63 cases, or 46% of the time. One of the 18 participants had no stimulated recall record because of technical difficulty. All 18 participants had a valid heat map record. Table 1 shows a noticing event (HM+) for the targeted items (after the recast) during the chat interaction in 90 of the 127 cases, or 71% of the time. No noticing event was apparent via heat map in 37 cases, or 29% of the time.

Table 1. *Noticing Events Across the Measures Stimulated Recall and Heat Map*

	Stimulated Recall		Heat Map	
	-/0	+/1	-/0	+/1
n (%)	63 (46%)	74 (54%)	37 (29%)	90 (71%)
Valid cases	137		127	

Table 2 shows the overall proportion of recast items scored "correct" on the posttests. Of all of the non-target-like items that were recast by the researcher and for which we have immediate and delayed posttest data, about 59% and 41% were scored as "correct" on the immediate and delayed posttests respectively.

Table 2. *Overall Immediate and Delayed Post-test Scores for all Valid Cases of Recasts*

	Total number of recasts	Number scored correct	Overall % correct
Immediate Posttest	121	71	59%
Delayed Posttest	85	35	41%

Research Question 1

What is the relationship between the individual measures of noticing (heat map and stimulated recall) and posttest scores for the targeted recast items?

Tables 3 and 4 below show the descriptive data and overall immediate and delayed posttest performance for recast items (noticed and not noticed) based on each of the measures of noticing employed. One notices that the stimulated recall and heat map records are quite similar to one another in terms of the number/percentage of recasts with positive and negative noticing records as well as the respective posttest scores for those same items.

Table 3. *Posttest Scores on Individual Target Items Based on Stimulated Recall Measure*

Noticing record	Immediate posttest				Delayed posttest			
	n	(%)	Mean score	*SD*	*n*	(%)	Mean score	*SD*
SR-	46	(43%)	.15	.36	43	(52%)	.15	.35
SR+	62	(57%)	.87	.34	40	(48%)	.73	.45
Total	108	(100%)	.56	.50	83	(100%)	.42	.50

Taken together, the descriptive data suggest that the heat map and the stimulated recall measures, though strong predictors of IPT success (IPT = 1), and to a certain extent DPT success, were very strong indicators of what learners likely failed to notice given the very low posttest means for those recasts coded as HM- and SR- respectively.

Table 4. *Posttest Scores on Individual Target Items Based on Heat Map Measure*

Noticing record	Immediate posttest				Delayed posttest			
	n	(%)	Mean score	*SD*	*n*	(%)	Mean score	*SD*
HM-	33	(32%)	.09	.29	27	(40%)	.15	.36
HM+	69	(68%)	.81	.39	41	(60%)	.66	.48
Total	102	(100%)	.58	.50	68	(100%)	.46	.50

From these tables we see a considerably higher percentage of recasts with a positive noticing event record in the heat map condition (about 68%) relative to the stimulated recall condition (about 57%). However, the mean IPT score of noticed items is slightly higher in the stimulated recall condition and more noticeably so in the DPT. Likewise, the mean IPT score on the recast items not noticed (HM-) is considerably lower in the heat map condition than in stimulated recall. This suggests two things: first, it suggests that the heat map measure is picking up on recast items that are attended to at some level, but this engagement often does not translate into the ability to productively use (in writing) the target-like form of the recast item a short time later. It also suggests that if the heat map record shows no evidence of noticing (HM-), then it reasonably certain that the item was not noticed at any level. In contrast, evidence of noticing based on the stimulated recall measure is a more solid indicator that learners have noticed a targeted item; at least at a level that allows target-like production a short time later. When comparing heat map and stimulated recall, the higher immediate posttest scores for items in the stimulated recall

condition that were not noticed (SR-) suggest that learners perhaps did notice many of these items at some level, but were not able to articulate this during the stimulated recall session.

Because participants received an unequal number of recasts from the interlocutor, the 156 observations were not independent of one another and were grouped by participant. This guided the statistical analysis chosen as most of the more commonly used statistical analysis procedures assume an independence of observations. In order to take this non-independence into account, a Generalized Estimating Equation for Logistic regression (GEE logit in SAS) was used (with exchangeable working matrix) (hereafter GEE logit). The GEE logit estimates the same model as the standard logistic regression but GEE logit allows for dependence within clusters (participants). This analysis was conducted for each measure of noticing (stimulated recall and heat map) for both immediate (IPT) and delayed (DPT) posttests.[17]

Table 5 shows the likelihood that a particular recast item would be scored correct on the posttest when it showed a positive noticing event (SR+, HM+) by the two measures of noticing.[18] For example, we can see that during the stimulated recall session, when learners indicated that they had noticed a targeted item in a recast (SR+), then the chances that this participant would get that same targeted item correct on the Immediate Posttest was 38.43 times greater than in the SR- condition ($p < .0001$). Thus, Table 5 suggests that both measures of noticing, when considered individually, seem to be generally favorable predictors of immediate posttest success with heat map being the stronger of the two. In terms of delayed posttest success the two measures of noticing seem to be much more similar in predicting the probability of success. The DPT value for stimulated recall shows that when participants have indicated SR+ for an item, that they are over 15 times more likely to score the item correct on the posttest than when they indicated SR-. This is a bit higher than the same statistic for heat maps, but may be (partially) explained by the fact that whereas the IPT occurs before the stimulated recall session, the DPT occurs afterward, allowing for a second exposure.

Table 5. *Relationship Between Evidence of Noticing and Posttest Success*

Measure of noticing	Post test	Observations used	Exp(log (Odds Ratio))	Standard error	*p*-value	C.I.*
Stimulated recall	IPT	108	38.43	1.72	<.0001	13.27 – 111.39
	DPT	83	15.86	1.69	<.0001	5.70 – 44.21
Heat map	IPT	102	48.52	1.81	<.0001	15.18 – 154.47
	DPT	68	10.84	2.47	.008	1.85 – 63.62

Notes. * alpha = .05. IPT = Immediate posttest. DPT = Delayed posttest.

Research Question 2

What is the relationship between heat map and stimulated recall to one another when used as measures of noticing?

To help answer Research Question 2, which focused on the possible relationship between heat map and stimulated recall, the same type of GEE logit was used as in the analysis reported in Table 5 above.

Table 6. *Probability of Co-occurrence of Measures of Noticing*

Correlation pair	Observations used	Exp (log(Odds Ratio))	Standard error	*p*-value	C.I.* of odds ratio
Heat map Stimulated recall	114	11.45	1.64	<.0001	4.36 – 30.08

Note. * alpha = .05

Table 6 reflects the likelihood that stimulated recall will indicate noticing (SR+) when heat map also indicates noticing (HM+). That is to say, when the heat map indicated a noticing event (HM+), it was 11.45 times as likely that the stimulated recall measure would also indicate noticing (SR+) than when the heat map measure did not indicate a noticing event (HM-).[19]

Turning our attention now to how well the combination of measures predict posttest success, Table 7 below shows that when considering heat map in conjunction with stimulated recall, both measures add positively and significantly to the model predicting IPT success. In this case when both measures are included in the model simultaneously, stimulated recall is a slightly stronger predictor of IPT success than heat map. In terms of the DPT, heat map fails to add any predictive weight to determining posttest success, whereas the stimulated recall measure continues to do so. Again, this is likely related to the fact that in the DPT participants had the benefit of a second exposure to the corrective feedback.

Table 7. *Heat Map and Stimulated Recall as Predictors of Posttest Success*

Combinations of measure of noticing	Post test	Observations used	Estimate	Standard error	*p*-value	C.I.*
Heat map	IPT	93	22.11	2.16	<.0001	4.88 – 99.98
Stim. recall			29.05	2.32	<.0001	5.58 –151.41
Heat map	DPT	66	6.60	3.29	.08 (ns)	-1.24 – 52.80
Stim. recall			4.91	1.54	.0003	2.09 – 11.51

Note. * alpha = .05

Table 8 shows the same data as in Table 7, but in a slightly different way. In Table 8 we see that posttest scores for each recast item for which there exists a valid stimulated recall and heat map record (IPT *n* = 93; DPT *n* = 66). There are four possible combinations for the noticing record. These are listed under the heading *measure of noticing*.

Table 8. [20] *Posttest Scores for Individual Target Items by Heat Map (HM) Stimulated Recall (SR) Record*

	Measure of noticing		IPT score (*n* = 93)		DPT score (*n* = 66)	
	HM	SR	0	1	0	1
Row 1	0	0	24 (.92)	2 (.08)	20 (.91)	2 (.09)
Row 2		1	6 (.86)	1 (.14)	3 (.60)	2 (.40)
Row 3	1	0	8 (.89)	1 (.11)	4 (.57)	3 (.43)
Row 4		1	2 (.04)	49 (.96)	8 (.25)	24 (.75)

As expected, the first row shows that when neither stimulated recall nor heat map indicate noticing, then those items are typically scored incorrect on the IPT. To be clear, this means that learners in these cases produced the same faulty formulation on the IPT that they used in the chat interaction even after receiving a target-like recast from the interlocutor. Also as expected, row four shows an overwhelming score of *correct* when both the stimulated recall and heat map records show evidence of noticing. Most interesting are rows two and three. The second row shows seven instances of the combination HM-/SR+ (IPT score). The data here are in the direction expected. A HM-record suggests that these items were not viewed by the learner and, therefore, were not available for further processing. I suggest that the six cases of *incorrect* despite the SR+ score indicate that learners misreported what they noticed during the task. That is to say, that they actually noticed the recast for the first time during the stimulated recall session, but for some reason reported having noticed it during the chat interaction. Another possible but unlikely explanation is that in these cases they did in fact notice the targeted item (and the eye tracker did not pick

up on these fixations) and their incorrect score on the IPT was simply a performance error.

The third row shows nine instances of the combination HM+/SR- for the IPT.[21] In these cases learners gave no indication in the stimulated recall session of having noticed the recast item during the chat whereas their heat map record suggests there was some heightened degree of attention to it. I will argue in the next section that these cases present evidence of learners attending to the targeted items in the input at some level (beyond that required for simply reading it, in which case we would expect a heat map color of purple), but not at the level of noticing with awareness. The percentage of such cases scored incorrect to correct was 89% to 11% respectively with only one instance of a positive IPT score. The data from the DPT shows essentially the same patterns, but we must be cautious in interpreting the DPT data since learners in this case had the benefit of re-exposure to the target items and some degree of explicit attention to form during the stimulated recall sessions as well as additional time (one week delay).

In summary, both measures of noticing, when considered independently, are positive predictors of IPT and DPT success. Further, where one measure suggests a noticing event, then there is a high likelihood that the other measure will indicate a noticing event as well. Stimulated recall reflects a slightly less powerful indication of IPT success (indeed, perhaps negligibly so, see Table 5) arguably because of memory decay which occurs between the chat interaction and the IPT. In contrast, the slightly more powerful indication of DPT success afforded by the stimulated recall measure may suggest one of two things. First, this measure consistently reflects a level of noticing with at least a low level of awareness whereas the heat map measure may not. Second, since the stimulated recall measure occurred after the task had been completed, whereas the heat map measure is recorded online during the task, we must assume that this fact could have had an effect on the DPT scores. That is to say, it may be that the DPT scores for the stimulated recall condition are high (partially) because of the stimulated recall procedure, which essentially becomes part of the treatment. Thus, unless we are to discuss the pedagogical value of stimulated recalls, then it is best to proceed cautiously with our interpretations of any apparent differences in the DPT scores of the two measures and stress, rather, their similarity.

Building on the results above that show the likelihood of posttest success based on noticing events, the next research question sought to examine the possible relationship between the linguistic category of the recast and noticing. Though because of the modest sample size these results are not explicitly tied to the posttest measures, the assumption here is that those recasts showing a noticing event would result in similar posttest patterns as those found in Research Questions 1 and 2.

Research Question 3

What is the relationship between the linguistic category of the recast target and noticing?

In this section I will examine the relationship between linguistic category on the noticing of recasts as measured by heat maps and stimulated recall.

Since posttest success was often used above as the major dependent variable in evaluating the viability of the two measures of noticing, it was important to eliminate the possibility that linguistic category itself (rather than noticing) had an effect on the posttest gain scores. That is to say, it may be reasonable to think that learners simply score higher on target items of one linguistic category over the others for some reason unrelated to noticing.

Since each recast was coded for *linguistic category*, a GEE logit was used as in Research Questions 1 and 2 above. This analysis showed no significant main effect for linguistic category for either the immediate posttest or delayed posttest scores. Table 9 below shows the descriptive data for the immediate and delayed posttest scores respectively across the three linguistic categories. These data suggest that posttest scores did not vary (statistically) as a function of linguistic category, though for both tests we do see markedly lower scores for syntax (and also a very low *n*).

Table 9. *Overall Posttest Success across Linguistic Category*

Linguistic Category of Recast	Immediate posttest			Delayed posttest		
	n	Mean	*SD*	*n*	Mean	*SD*
Morphology	49	0.53	0.50	38	0.37	0.49
Semantics	67	0.64	0.48	40	0.48	0.51
Syntax	5	0.40	0.55	7	0.29	0.49
Total	121	0.59	0.49	85	0.41	0.49

Note. n =cases where there were observations for the respective categories and where there was a posttest score for that item.

Table 10 shows the descriptive data by linguistic category across the measures of noticing. In order to answer this research question a GEE logit was conducted for the variable linguistic category for the two measures of noticing.

Table 10. Linguistic Category of Recast by Measure of Noticing

Linguistic Category	Measure of noticing			
	Heat map		Stimulated recall	
	No evidence of noticing	Evidence of noticing	No evidence of noticing	Evidence of noticing
Morphology	20 (37.5%)	36 (62.5%)	33 (57.9%)	24 (42.1%)
Semantics	15 (23.1%)	50 (76.9%)	27 (38.0%)	44 (62.0%)
Syntax	2 (33.3%)	4 (66.7%)	3 (33.3%)	6 (66.7%)
Total	37	90	63	74

Results for the heat map measure showed no significant relationship between linguistic category and noticing. However, there was a significant main effect for stimulated recall. The odds of semantic recasts resulting in a noticing event in the stimulated recall measure was 2.12 times higher than for morphological recasts ($p = 0.006$), 95% (exp) CI (1.24, 3.64). Likewise, the odds of syntactic recasts resulting in noticing in the stimulated recall measure was 3.73 times higher than for morphological recasts ($p = 0.0002$), 95% (exp) CI (1.87, 7.47). No significant relationship was found between semantic recasts and syntactic recasts. This suggests that when using the stimulated recall methodology to assess learners' noticing of recasts, semantic and syntactic recasts were two and three times (respectively) more likely to be noticed than were morphological recasts. The heat map measure of noticing (though generally in the same direction) did not show this pattern as clearly.

In order to determine if the measures of noticing differed in how they reflected the extent to which recasts from each linguistic category were noticed, a series of paired sample *t*-tests were run (see Tables 11 and 12). For morphology, the data show that the heat map measure was more likely to yield a positive noticing event, whereas no significant difference was found for semantics. This supports the notion that

Table 11. *Percentage of Target Recasts Showing Positive Noticing Events across the Linguistic Category of Morphology*

Morphological Recasts	*n*	Mean	*SD*	$\overline{X1}-\overline{X2}$	*SD of diff.*	C.I.	*t*	*df*	*p*	*d*
Stimulated recall Heat map	14	.47 .62	.31 .43	-.15	.23	-.28 – .02	-2.41	13	.031	.65

explicit written SCMC recasts targeting morphology, though attended to at some level, are more difficult than semantic recasts to notice at any useful level as measured by stimulated recall. Further analysis of syntactic recasts was not advisable due to the extremely low number of possible paired comparisons.

Table 12. *Percentage of Target Recasts Showing Positive Noticing Events Across the Linguistic Category of Semantics*

Semantic Recasts	*n*	Mean	*SD*	$\overline{X1}-\overline{X2}$	SD of diff.	C.I.	*t*	*df*	*p*	*d*
Stimulated recall Heat map	17	.70 .77	.27 .28	-.07	.25	-.20 – .06	-1.13	16	.276	.28

Notes. For Tables 11 and 12, *n* refers to the number of participants with valid data for each paired comparison. The effect size measure Cohen's d for paired groups was employed using the formula Cohen's d = mean difference between the pairs/standard deviation of the difference (see Kotrlik & Williams, 2003).

DISCUSSION

When considered independently from one another, the two measure of noticing employed in this study appear to be favorable predictors of noticing, with the heat map measure being slightly more discriminating in its ability to predict the probability of IPT success than stimulated recall. This difference, however, is a matter of degree. Discriminating in the sense that when the heat map record indicates noticing (HM+) then the likelihood that the relevant item is scored "correct" (IPT = 1) is over 48 times higher than in the HM- condition. This is in contrast to the 38 times higher for stimulated recall.

The theoretical value of employing eye tracking in explorations of learner attention is that it adds a new and powerful methodological dimension in exploring constructs associated with attention and noticing and their respective roles in SLA. Though the current study was not designed to specifically examine possible levels of awareness with respect to corrective feedback, the modest data presented here might serve as a springboard for future more in-depth investigations into this area. For example, we may interpret Table 8 in the following way: Assuming that the eye tracker was properly calibrated, rows 1 and 2 (HM = 0) seem to show no evidence of detection since no significant eye gaze was registered over the targeted items. We may say, then, that there seems to have been no registration of the stimuli at all. Of course, there are three cases of IPT success for these items; however, over 90% of these cases led to an IPT score of zero. In rows 3 and 4 we have evidence of the cognitive registration of the stimuli (Tomlin & Villa, 1994 at some level (detection). Schmidt (2001) suggests that it is necessary to distinguish between 1. detection without awareness and 2. detection within focal attention accompanied by awareness, that is, conscious perception or noticing. Making this distinction, as he further points out, is not an easy task. For example, how can we know whether some stimulus (or a feature of it) has been attended? And how can we know whether some stimulus (or a feature of it) has been noticed (Schmidt, 2001)? I argue that eye tracking data may help us make this distinction and we can do this in a pedagogically relevant way. In terms of the present data, items coded in Table 8 as HM = 1, SR = 0 may be viewed as having been attended to at some level but not noticed (detected without awareness). Viewed another way, it could be argued that these items were detected but did not reach a sufficient threshold of rehearsal in short-term memory (Robinson, 1995). Those items coded as HM = 1, SR = 1 may be viewed as evidence of detection with awareness (noticed), at least a low level of awareness (see Leow, 2000). In both cases the IPT scores confirm this view.

The results in terms of which linguistic category of recast was noticed more readily are mixed, and vary as a function of the measure employed. There was no significant relationship found for linguistic category

and noticing in the heat map condition. For the stimulated recall condition, however, learners were much more likely to report having noticed semantic recasts than morphological recasts. Syntactic recasts also hold the same advantage over morphological recasts, though the extremely low numbers of syntactic recasts in the data make any real claims in this regard problematic. This is consistent with previous research by Mackey et al. (2000) and Nabei and Swain (2002) who had positive findings for noticing lexical feedback over other types. This evidence is also compatible with an input processing theoretical perspective (Van Patten, 2007), which argues that learners prefer to process information lexically over other features in the input.

CONCLUSION

This study offers a foray into the use of eye tracking technology as a measure of noticing corrective feedback. Though measuring what learners attend to in the input in any precise way remains a difficult challenge for researchers, the use of eye gaze data seems to be potentially valuable in helping to determine which features of the input are likely to be noticed and which are not since we can see precisely what learners view and arguably attend to. Used in conjunction with stimulated recall, the eye tracking methodology may help researchers untangle the lower levels of the noticing continuum. This, of course, remains to be seen. Finally, eye tracking may provide an avenue whereby many of the argued benefits of text-based SCMC, such as its facility to enhance the salience of input, can be put to the test.

One practical conclusion of the study is that stimulated recall and heat maps are highly (and positively) correlated with one another and in cases where there is a positive record of both, then the likelihood that the targeted item has reached a higher level of learner awareness (ultimately confirmed by their IPT scores for that same item) is extremely high. For SLA researches this is good news given the pervasiveness of stimulated recall as a methodological technique.

APPENDIX. Delayed Posttest Directions

Delayed Posttest directions
Greetings!
Thank you for taking part in the English study the other day!

In an email that I will send a little later you will receive a copy of what
you wrote the other day in the English study chat session. I have removed my
half of the dialog so you will only see what YOU wrote.

When you have a minute over the next day or so I would like you to look over
these sentences you wrote. Please do NOT use any outside sources.

Please don't open the second email as I am timing how long it takes for each
student to complete this activity based on when they SEND me the completed
email. That is, when you actually open the email I will receive a message
that tells me that you have viewed or opened it. I will give you one minute
to complete each sentence. This means that if I send you 15 sentences, then
you should complete the activity and send it back to me within 15 minutes.
In the second email "subject line" I will write how many sentences you have
to look at. Please finish the activity once you begin since you cannot
return to it later.

Directions:
Simply decide if the sentences you wrote are correct or not as they are
written. That is, are they grammatical as they are written here or are there
one or more errors in the sentence? If the sentence is OK, then simply write
OK under it. If there is an error, please write "No" and then re-write the
sentence and FIX the error. Please and email me this back as soon as you
can.

Thank you!

Here is an example:

1.   [12:00:01] <Eye Tech> The cow jump over the moon.

Your answer: No   The cow jumped over the moon.

Researcher's Comment: See the past tense "-ed" added after "jump" from line 1?

2.   [12:00:01] <Eye Tech> The blue guy climbed the tree.

Your answer: OK

Researcher's Comment: Line 2 is OK the way it is.

 Please complete these:
[SENTENCES VARY BY PARTICIPANT]

NOTES

1. However, a failure to notice a target cannot be inferred by a failure to verbalize something.

2. It is possible that these positive findings could be attributed to the narrow discourse contexts often
present in experimental environments. That is to say, recasts in these (largely) experimental studies are
often isolated from other feedback strategies and are frequently provided intensively by the interlocutor.
Oliver (1995), for example, suggests that the interpretation of recasts is greatly influenced by the
discourse context.

3. Though the learner internal variables of proficiency level and working memory capacity are viewed as
important factors in explaining what learners notice in the linguistic input, a discussion of the respective
roles played by these two variables is beyond the scope of this paper.

4. However, some research shows recasts to be effective for other aspects of language as well (Doughty &
Varela, 1998; Sagarra, 2007; Sheen, 2006).

5. A contingent recast in this context referred to recasts that immediately followed the trigger turn and
appeared just below the trigger turn on the screen.

6. However, regressive saccades (i.e., backtracking eye-movements) can act as a measure of processing
difficulty during encoding (Rayner & Pollatsek, 1989). Indeed, between 10–15% of the time readers
perform regressive saccadic movements to return to material that has been already read (Dussias, 2010).

In this study, we will be concerning ourselves with fixations only.

7. One participant submitted a corrupted pre-test file, so has no pre-test record. An comparison of this participant's self-reported TOEFL score put this student in the lower half of those reporting TOEFL scores. Note that not all participants reported having taken the TOEFL test.

8. Skewness and Kurtosis were -.837 and -.688 respectively. A value between 0 and +/- 1 was viewed as adequate for this comparison (*n* = 17).

9. The design initially asked students to think aloud during the task as well. However, the data showed that only two of the participants engaged in think aloud to a sufficient degree. Several participants had no think aloud record at all. Thus, the decision was made to omit this aspect of the data from the study since a comparison of these data with that of the stimulated recall and eye gaze records would be problematic given the relatively small *n* in this study.

10. The author acknowledges that this feature of the research design moves it away from a more natural setting since such training primes participants to notice such feedback in the input.

11. Komínková (2008) found no significant differences in the stability of the calibration between head-mount and remote eye tracking systems examined. In that study the head-mount gaze position accuracy was 95–99% at 2.5–3° viewing angle and for the remote position accuracy was 95–99% at 2–3° viewing angle.

12. http://en.wikipedia.org/wiki/Plonsters

13. The following is a segment of the chat transcript that illustrates the nature of the interaction.

    [13:32:48] <Learner> There is an orange guy who feels so sleepy
    [13:32:52] <Interlocutor> ok
    [13:33:03] <Learner> He goes to bed and take a nap
    [13:33:23] <Interlocutor> OK, so he goes to bed and takes a nap -- good!
    [13:33:40] <Learner> Two of his friends come to see him
    [13:33:52] <Interlocutor> OK, that's nice of them!
    [13:34:05] <Learner> They want to ask him go out
    [13:34:35] <Interlocutor> OK, they want to ask him to get up?
    [13:34:47] <Learner> yes
    [13:34:51] <Interlocutor> ok
    [13:35:03] <Learner> They try to find a way to wake him up
    [13:35:16] <Interlocutor> Alright, what do they do?
    [13:35:50] <Learner> First, one of his friends transforms himself into an alarm clock
    [13:36:06] <Interlocutor> oh, OK -- why is that?

14. Of course, with this design there is no way to realistically control for external variables such as help or the use of outside materials. As such, the design is sensitive to threats to internal validity. Nevertheless, since this is a within groups design, the potential impact of such external variables is viewed here as a limitation to the DPT results, albeit one that applies equally to all participants. Further, since learners were given one minute on average to complete each question on the DPT, the likelihood of their effectively getting "help" on any one question is minimized.

15. The time that the DPT was opened was recorded automatically via the *return receipt* function in Microsoft Entourage.

16. Smith (2010) sets 500ms as the threshold for coding a linguistic unit as noticed.

17. The exchangeable covariance structure was chosen for several reasons. First, since the observations on each participant will be correlated, it seems reasonable to assume that the correlation is same for

any two observations. This option also minimizes the number of parameters that need to be estimated. The exchangeable matrix requires only calculation of two variance components. Finally, there is some evidence from the literature that GEE is robust to the structure of covariance matrix chosen (see Park and Shin, 1999 for a discussion of the use of working correlation matrices in the GEE approach for longitudinal data).

18. Cases where a targeted item was recast during the task but where there was no attempted use apparent from the immediate posttest were omitted from the analysis. All items recast were included in the delayed posttest.

19. Table 6 reflects the GEE analysis when the variable Heat Map is entered into the statistical model first and Stimulated Recall second. When Stimulated Recall is entered into the model first, we have a similar finding (Exp(log(Odds Ratio)) = 9.10; *SE* 1.60; $p < .0001$; C.I. of odds ratio 3.62–22.89).

20. Tables 1–8 do not distinguish between semantic, morphological, and syntactic targets. Rather they reflect the number of recasts in general. Data for the linguistic category of the recasts is reflected in Tables 9–12.

21. There were a few ambiguous cases in this category, which were removed from the data for this analysis ($n = 5$). For example, in one case a possessive pronoun was omitted in an obligatory case (its). The recast provided the required pronoun, in the IPT, however, the learner used the masculine possessive pronoun his rather than the neuter possessive pronoun. Though this could be viewed as a positive change, the threshold for scoring an item correct on the posttests was a target-like change that reflected the exact language from the recast.

ACKNOWLEDGEMENTS

REFERENCES

Alanen, R. (1995). Input enhancement and rule presentation in second language acquisition. In R. Schmidt (Ed.), *Attention and awareness in foreign language learning and teaching* (pp. 259–302). Honolulu: University of Hawai'i Press.

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

Braidi, S. M. (2002). Reexamining the role of recasts in native-speaker/nonnative-speaker interactions. *Language Learning, 52*(1), 1–42.

Brown, R., & U. Bellugi (1964). Three processes in the child's acquisition of syntax. *Harvard Educational Review, 34,* 133–151.

Carroll, S., & Swain, M. (1993). Explicit and implicit negative feedback: An empirical study of the learning of linguistic generalizations. *Studies in Second Language Acquisition, 15,* 357–386.

Chaudron, C. (1977). A descriptive model of discourse in the corrective treatment of learners' errors. *Language Learning, 27,* 29–46.

Dasse-Askildson, V. (2008). How learners' affective variables impact their perception of recasts in the acquisition of grammatical gender in L2 French. *Arizona Working Papers on SLA and Teaching, 15,* 1–

35. Retrieved from http://w3.coh.arizona.edu/awp/AWP15/AWP15%5BDasse%5D.pdf

Doughty, C. (1994). Fine-tuning of feedback by competent speakers to language learners. In J. Alatis (Ed.), *Strategic interaction and language acquisition: Theory, practice, and research.* GURT 1993 (pp. 96–108). Washington: Georgetown University Press.

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.

Dussias, P. (2010). Uses of eye-tracking data in second language sentence processing research. *Annual Review of Applied Linguistics, 30,* 149–166.

Egi, T. (2007). Interpreting recasts as linguistic evidence: The roles of length and degree of change. *Studies in Second Language Acquisition, 29*, 511–537.

Egi, T. (2010). Uptake, Modified Output, and Learner Perceptions of Recasts: Learner Responses as Language Awareness. *The Modern Language Journal, 94,* 1–21.

Ellis, R., & Yuan, F. (2004). The effects of planning on fluency, complexity and accuracy in second language narrative writing. *Studies in Second Language Acquisition, 26,* 59–84.

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

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.

Færch, C., & Kasper, G. (Eds). (1987). *Introspection in second language research.* Clevedon, UK: Multilingual Matters.

Fanselow, J. F. (1977). The treatment of error in oral work. *Foreign Language Annals, 10,* 583–593.

Gass, S. M. (1988). Integrating research areas: A framework for second language studies. *Applied Linguistics, 9,* 198–217.

Gass, S. M. (1997). *Input, interaction, and the second language learner.* Mahwah, NJ: Erlbaum.

Gass, S. M., & Mackey, A. (2000). *Stimulated recall methodology in second language research.* Mahwah, NJ: Lawrence Erlbaum Associates.

Gass, S. M., & Varonis, E. M. (1994). Input, interaction, and second language production. *Studies in Second Language Acquisition, 16, 283–302.*

Godfroid, A., Housen, A., & Boers, F. (2010). A procedure for testing the noticing hypothesis in the context of vocabulary acquisition. In M. Pütz & L. Sicola (Eds.), *Cognitive processing in second language acquisition* (pp. 169–197). Amsterdam: John Benjamins.

Goo, J., & Mackey, A. (in press). The case against the case against recasts. *Studies in Second Language Acquisition.*

Izumi, S. (2003). Comprehension and production processes in second language learning: In search of the psycholinguistic rationale of the output hypothesis. *Applied Linguistics, 24,* 168–196.

Jacob, R. J. K., & Karn, K. S. (2003). Eye tracking in Human-Computer Interaction and usability research: Ready to deliver the promises, In J. Hyönä, R. Radach, & H. Deubel (Eds.), *The mind's eye: Cognitive and applied aspects of eye movement research* (pp. 573–605). Amsterdam: Elsevier.

Just, M. A., & Carpenter, P. A. (1976). Eye fixations and cognitive processes. *Cognitive Psychology, 8,* 441–480.

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, *87*, 329–354.

Kim, J. H., & Han, Z-H. (2007). Recasts in communicative EFL classes: Do teacher intent and learner interpretation overlap? In A. Mackey (Ed.), *Conversational Interaction in second language acquisition: A series of empirical studies* (pp. 269–297). Oxford: Oxford University Press.

Komínková, B. (2008). *Comparison of two eye tracking devices used on printed images.* Unpublished master's thesis. Department of Graphic Arts and Photophysics, University of Pardubice, Poland.

Kotrlik, J. W., & Williams, H. A. (2003). The incorporation of effect size in information technology, learning, and performance research. *Information Technology, Learning, and Performance Journal, 21*(1), 1–7.

Kuhn, J. (2012). The noticing of correct and incorrect forms in lengthier texts: An ESL eye-tracking investigation. In G. Kessler, A. Oscoz, & I. Elola (Eds.), *Technology across writing contexts and tasks* (pp. 237–253). San Marcos, TX: CALICO.

Lai, C., Fei, F., & Roots, R. (2008). The Contingency of Recasts and Noticing. *CALICO Journal, 26,* 70–90.

Leeman, J. (2003). Recasts and second language development: Beyond negative evidence. *Studies in Second Language Acquisition, 25,* 37–63.

Leow, R. P. (1997). Attention, awareness, and foreign language behavior. *Language Learning, 47,* 467–506.

Leow, R. P. (2000). A study of the role of awareness in foreign language behavior: Aware vs. unaware learners. *Studies in Second Language Acquisition, 22,* 557–584.

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

Liu, Y. (2008). The effects of error feedback in second language writing. Arizona *Working Papers in SLA & Teaching, 15,* 65–79.

Loewen, S., & Erlam, R. (2006). Corrective feedback in the chatroom: An experimental study. *Computer Assisted Language Learning, 19,* 1–14.

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

Long, M. H., Inagaki, S., & Ortega, L. (1998). The role of implicit negative feedback in SLA: Models and recasts in Japanese and Spanish. *The Modern Language Journal, 82*, 357–371.

Long, M. H. (1996). The role of the linguistic environment in second language acquisition. In W. C. Ritchie & T. K. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 413–468). San Diego, CA: Academic Press.

Long, M. H. (2007). *Problems in SLA*. Mahwah, NJ: 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, 19,* 37–66.

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

Mackey, A. (2006). Feedback, noticing and instructed second language learning. *Applied Linguistics, 27,* 405–430.

Mackey, A., Gass, S., & McDonough, K. (2000). How do learners perceive interactional feedback? *Studies in Second Language Acquisition, 22*, 471–497.

Mackey, A., & Goo, J. (2007). Interaction research in SLA: A meta-analysis and research synthesis. In A. Mackey (Ed.), *Conversational interaction in second language acquisition: A collection of empirical studies* (pp. 407–451). New York: Oxford University Press.

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

Mackey, A., Philp, J., Egi, T., Fujii, A., & Tatsumi, T. (2002). Individual differences in working memory, noticing of interactional feedback and L2 development In P. Robinson (Ed.), *Individual differences and instructed language learning* (pp. 181–208). Philadelphia: John Benjamins.

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

Morris, F. A., & Tarone, E. E. (2003). Impact of classroom dynamics on the effectiveness of recasts in second language acquisition. *Language Learning, 53,* 325–368.

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*, 43–62.

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

O'Rourke, B. (2008). The other C in CMC: What alternative data sources can tell us about textbased synchronous computer mediated communication and language learning. *Computer Assisted Language Learning, 21,* 227–251.

O'Rourke, B. (in press). Using eye tracking to investigate gaze behaviour in synchronous computer-mediated communication for language learning. In M. Dooly & R. O'Dowd (Eds.), *Researching online interaction and exchange in foreign language education: Methods and issues*. Frankfurt am Main: Peter Lang.

Ohta, A. (2000). Rethinking recasts: A learner-centered examination of corrective feedback in the Japanese classroom. In J. K. Hall & L. Verplaetse (Eds.), *The construction of second and foreign language learning through classroom interaction* (pp. 47–71). Mahwah, NJ: Erlbaum.

Oliver, R. (1995). Negative feedback in child NS-NNS conversation. *Studies in Second Language Acquisition,* 17, 459–481.

Oliver, R., & Mackey, A. (2003). Interactional context and feedback in child ESL classrooms. *Modern Language Journal, 87*(4), 519–533.

Ortega, L., & Long, M. H. (1997). The effects of models and recasts on the acquisition of object topicalization and adverb placement by adult learners of Spanish. *Spanish Applied Linguistics, 1,* 65–86.

Park, T., & Shin, D-Y. (1999). On the use of working correlation matrices in the GEE approach for longitudinal data. *Communications in Statistics - Simulation and Computation, 28,* 1011–1029.

Pellettieri, J. (1999). *Why-talk? Investigating the role of task-based interaction through synchronous network-based communication among classroom learners of Spanish.* Unpublished doctoral dissertation. College of Letters and Sciences, Department of Spanish, University of California, Davis.

Philp, J. (1999). *Interaction, noticing, and second language acquisition: An examination of learners' noticing of recasts in task-based interaction.* Unpublished doctoral dissertation. School of Letters, Arts, and Media, Department of Linguistics, University of Tasmania, Australia.

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.

Poole, A. & Ball, L. J. (2006). Eye Tracking in Human-Computer Interaction and Usability Research: Current Status and Future Prospects. In C. Ghaoui, (Ed.). *Encyclopedia of Human Computer Interaction* (pp. 211–219). Hershey, PA: Idea Group

Poole, A., Ball, L. J., & Phillips, P. (2004). In search of salience: A response time and eye movement analysis of bookmark recognition. In S. Fincher, P. Markopolous, D. Moore, & R. Ruddle (Eds.), *People and Computers XVIII-Design for Life: Proceedings of HCI 2004.* London: Springer-Verlag Ltd.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 12,* 372–422.

Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology,* in press.

Rayner, K., & Pollatsek, A. (1989). *The psychology of reading.* Hillsdale, NJ: Erlbaum.

Révész, A. (2009). Task complexity, focus on form, and second language development. *Studies in Second Language Acquisition, 31,* 437–470.

Robinson, P. (1995). Attention, memory, and the noticing hypothesis. *Language Learning, 45,* 283–331.

Robinson, P. (1996). Learning simple and complex second language rules under implicit, incidental, rule-search and instructed conditions. *Studies in Second Language Acquisition, 18,* 27–67.

Rosa, E. M., & O'Neill, M. (1999). Explicitness, intake, and the issue of awareness: Another piece to the puzzle. *Studies in Second Language Acquisition, 21,* 511–556.

Sachs, R., & Suh, B-R. (2007). Textually enhanced recasts, learner awareness, and L2 outcomes in synchronous computer-mediated interaction. In A. Mackey (Ed.), *Conversational interaction in second language acquisition: A collection of empirical studies* (pp. 197–227). Oxford: Oxford University Press.

Sagarra, N. (2007). From CALL to face-to-face interaction: The effect of computer-delivered recasts and working memory on L2 development. In A. Mackey (Ed.), *Conversational interaction in second language acquisition: A series of empirical studies.* (pp. 229–248). Oxford: Oxford University Press.

Salaberry, M. R. (2000). L2 morphosyntactic development in text based computer mediated communication. *Computer Assisted Language Learning*, *13*, 5–27.

Sauro, S. (2009). Computer-mediated corrective feedback and the development of second language grammar. *Language Learning & Technology*, *13*(1), 96–120. Retrieved from http://llt.msu.edu/vol13num1/sauro.pdf

Sauro, S., & Smith, B. (2010). Investigating L2 Performance in Text Chat. *Applied Linguistics 31*(4), 554–577.

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

158.

Schmidt, R. (1995). Consciousness and foreign language learning: A tutorial on the role of attention and awareness in learning. In R.W. Schmidt (Ed.), *Attention and awareness in foreign language learning* (pp. 1–63). Honolulu, HI: University of Hawai'i, Second Language Teaching and Curriculum Center.

Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3–32). New York: Cambridge University Press.

Seedhouse, P. (1997). The case of the missing "no": The relationship between pedagogy and interaction. *Language Learning, 47,* 547–583.

Sheen, Y. (2004). Corrective feedback and learner uptake in communicative classrooms across instructional settings. *Language Teaching Research, 8,* 263–300.

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

Sheen, Y. (2008). Recasts, language anxiety, modified output and L2 learning. *Language Learning, 58,* 835–874.

Skehan, P., & Foster, P. (2005). Strategic and on-line planning: The influence of surprise information and task time on second language performance. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 193–216). Mahwah, NJ: Lawrence Erlbaum Associates.

Smith, B. (2004). Computer-mediated negotiated interaction and lexical acquisition. *Studies in Second Language Acquisition, 26,* 365–398.

Smith, B. (2005). The Relationship between Negotiated Interaction, Learner uptake, and lexical acquisition in task-based computer-mediated communication. *TESOL Quarterly, 39,* 33–58.

Smith, B. (2008). Methodological Hurdles in capturing CMC data: The case of the missing self-repair. *Language Learning & Technology, 12*(1)*,* 85–103. Retrieved from: http://llt.msu.edu/vol12num1/smith/ default.html

Smith, B. (2010). Employing eye-tracking technology in researching the effectiveness of recasts in CMC. In Francis M. Hult (ed.), *Directions and prospects for educational linguistics* (pp. 79–98). New York: Springer.

Smith, B., & Renaud, C. (in press). Eye tracking as a measure of noticing corrective feedback in computer-mediated instructor-student foreign language conferences. In K. McDonough and A. Mackey (Eds.), *Interaction in diverse educational settings*. Philadelphia: John Benjamins.

Swain, M. (2006). Languaging, agency and collaboration in advanced second language learning. In H. Byrnes (Ed.), *Advanced language learning: The contributions of Halliday and Vygotsky* (pp. 95-108). London, UK: Continuum.

Tarone, E., & Bigelow, M. (2007). Alphabetic print literacy and processing of oral corrective feedback in L2 interaction. In A. Mackey (Ed.), *Conversational interaction in second language acquisition: A series of empirical studies* (pp. 101–121). Oxford: Oxford University Press.

Tomlin, R. S., & Villa, V. (1994). Attention in cognitive science and SLA. *Studies in Second Language Acquisition, 16,* 183–203.

Trofimovich, P., Ammar, A., & Gatbonton, E. (2007). How effective are recasts? The role of attention, memory, and analytical ability. In A. Mackey (Ed.), *Conversational interaction in second language acquisition: A series of empirical studies* (pp.144–171). Oxford: Oxford University Press.

Truscott, J., & Sharwood Smith, M. (2011). Input, Intake, and Consciousness: The Question for a

Theoretical Foundation. *Studies in Second Language Acquisition, 33,* 497–528.

VanPatten, B. (2007). Input processing in adult second language acquisition. In B. VanPatten & J. Williams (Eds.) *Theories in second language acquisition: An introduction* (pp. 115–135). Mahwah, NJ: Erlbaum.

Williams, J. (1999). Memory, Attention, and Inductive Learning. *Studies in Second Language Acquisition, 21,* 1–48.

Yuan, F., & Ellis, R. (2003). The effects of pre-task planning and on-line planning on fluency, complexity and accuracy in L2 oral production. *Applied Linguistics, 24,* 1–28.