<table>
<thead>
<tr>
<th>Title</th>
<th>Second language research on recasts: A critical review in response to an ongoing debate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Author(s)</td>
<td>Chengsong Yang, Guangwei Hu and Lawrence Jun Zhang</td>
</tr>
<tr>
<td>Published by</td>
<td>Walter de Gruyter</td>
</tr>
</tbody>
</table>

This document may be used for private study or research purpose only. This document or any part of it may not be duplicated and/or distributed without permission of the copyright owner.

The Singapore Copyright Act applies to the use of this document.

Second Language Research on Recasts: A Critical Review in Response to an Ongoing Debate

Chengsong YANG
School of International Studies, Xi’an Jiaotong University (China); Faculty of Education, University of Auckland (New Zealand)

Guangwei HU
National Institute of Education, Nanyang Technological University (Singapore)

Lawrence Jun ZHANG*
Faculty of Education, University of Auckland (New Zealand)

Abstract

This paper reviews important second language research on recasts in response to a recent debate between Goo and Mackey (2013) and Oyster and Ranta (2013) in the journal Studies in Second Language Acquisition. It begins by clarifying the definition of recasts and then examines several controversial issues featuring prominently in the debate, including the theoretical underpinnings of existing research on recasts, mixed findings about their effectiveness in language acquisition, and methodological concerns in effectiveness research. The primary aim of this paper is to demarcate common ground and disagreement that are emerging in this debate and, based on a synthesis and critique of previous empirical studies, to identify directions for future research. It also explores pedagogical implications of recasts research for English-as-a-foreign-language contexts.

Keywords: recasts, corrective feedback, interactionism, second language acquisition

* Corresponding author: Tel: +64 9 6238899x48750. E-mail address: ljzhang@auckland.ac.nz (L. J. Zhang).
1. Introduction

In defending the positive roles of recasts in second language (L2) acquisition within an interactionist paradigm, Goo and Mackey (2013) outline the methodological flaws and interpretative problems of a number of previous experimental studies that pointed to the (relative) ineffectiveness of recasts. They contend that these studies were subject to five types of design issues that have posed a serious threat to their internal validity. These issues concern (a) lack of control over opportunities for modified output in comparative research; (b) unfair comparisons of recast treatments with alternative conditions that often conflated multiple treatments; (c) the differential moderating effect of form-focused instruction (FFI) on recast treatment and comparison conditions; (d) lack of control for the effects of prior knowledge; and (e) the potential contamination of the experimental designs by out-of-experiment exposure. In response, Lyster and Ranta (2013) counter-argue (a) that the jury is still out on the relationship between recasts and L2 acquisition; (b) that their views on uptake have been misrepresented by Goo and Mackey; (c) that studies comparing different types of corrective feedback (CF) are theoretically meaningful, practically oriented, and empirically valid; and (d) that Goo and Mackey’s methodological recommendations close down promising avenues for further research on recasts. Both sides of the debate raise important theoretical and methodological questions about the state of the art of L2 recasts research. Given the significance of these questions and related issues, we believe that it is important to zoom in and out of the debate to establish consensus, demarcate disagreements, and identify promising directions for future research on CF in general and recasts in particular.

To establish a basis for our critical review, we will start by examining the continual scholarly effort to define recasts and extracting consensual elements out of the multitude of existing definitions. We will then take an overview of the theoretical landscape of L2 research on recasts and argue that our understanding in this regard is still expanding. Next, we will turn to the foci of the aforementioned debate and address topics central to recasts research, namely, uptake as a construct, (relative) effects of recasts, methodological issues in recasts research, and directions for future research. In each section of the discussion, we will first summarize the opposing views held by the two sides of the debate and then identify possible common ground out of divergences. We will end this review with a discussion on what pedagogical implications this debate and, more broadly, recasts research as a whole, may have for English-as-a-foreign-language (EFL) contexts such as mainland China.

2. Toward a consensual definition

“Recast” is a term borrowed from first language (L1) research but narrowed semantically in second language (L2) studies (Nicholas, Lightbown, & Spada, 2001). Based on Nicholas et al.’s discussion and following Ellis and Sheen’s (2006, p. 580) collection of definitions, various definitions of recasts are gleaned from the L2 research literature and presented in
Table 1. Along with Yilmaz’s (2012) and Goo and Mackey’s (2013) newest formulations, two earlier approximations, that is, Chaudron’s (1977, p. 39) term of “repetition with change” and Fröhlich, Spada, and Allen’s (1985, p. 56) term of “paraphrase,” are also included.

Table 1. Definitions of Recasts in L2 Research

<table>
<thead>
<tr>
<th>Reference</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chaudron (1977, p. 39)</td>
<td>“Repetition with change” is a teacher response to learner error that “simply adds correction and continues to other topics.”</td>
</tr>
<tr>
<td>Fröhlich, Spada, &amp; Allen (1985, p. 56)</td>
<td>“Paraphrase” is “completion and/or reformulation of previous utterance/s.”</td>
</tr>
<tr>
<td>Long (1996, p. 434)</td>
<td>Recasts are “utterances that rephrase a child’s utterance by changing one or more sentence components (subject, verb, or object) while still referring to its central meaning.”</td>
</tr>
<tr>
<td>Lyster and Ranta (1997, p. 46)</td>
<td>“Recasts involve the teacher’s reformulation of all or part of a student’s utterance, minus the error” (italics original).</td>
</tr>
<tr>
<td>Braidi (2002, p. 20)</td>
<td>“A response was coded as a recast if it incorporated the content words of the immediately preceding incorrect NNS utterance, and also changed and corrected the utterance in some way (e.g., phonological, syntactic, morphological or lexical).”</td>
</tr>
<tr>
<td>Sheen (2006, p. 365)</td>
<td>“The teacher’s reformulation of all or part of a student’s utterance that contains at least one error within the context of a communicative activity in the classroom”</td>
</tr>
<tr>
<td>Long (2007, p. 77)</td>
<td>“A corrective recast may be defined as a reformulation of all or part of a learner’s immediately preceding utterance in which one or more nontargetlike (lexical, grammatical, etc.) items is/are replaced by the corresponding target language form(s), and where, throughout the exchange, the focus of the interlocutors is on meaning, not language as object.”</td>
</tr>
<tr>
<td>Nassaji (2009, p. 412)</td>
<td>“Recasts refer to feedback that reformulates a learner’s nontargetlike utterance into a targetlike one.”</td>
</tr>
<tr>
<td>Yilmaz (2012, p. 1135)</td>
<td>“Recasts provide targetlike reformulations of learners’ incorrect utterances in a subtle manner without directly indicating (1) whether the learner has made an error, (2) which part of the utterance is nontargetlike, or (3) whether the reformulated version is actually the targetlike version of the learner’s utterance.”</td>
</tr>
<tr>
<td>Goo &amp; Mackey (2013, p. 129)</td>
<td>“Recasts are more targetlike versions of learners’ nontargetlike utterances.”</td>
</tr>
</tbody>
</table>

A scrutiny of the table reveals a tendency to associate recasts with communicative language teaching in some widely circulated definitions (e.g., Long, 1996, 2007; Sheen, 2006). For example, Long (2007) emphasizes that recasts are provided in a meaning-focused communicative context and direct learners’ attention to nontargetlike forms incidentally. As Ellis and Sheen (2006) note, however, Long’s two definitions in Table 1...
differ from each other in one crucial respect: that the more recent definition attempts to “exclude reformulations that refer to the central meaning of a learner utterance but that are clearly didactic (from the perspective of the person doing the recasting) rather than communicative (i.e., they do not constitute an attempt to solve a communication problem)” (p. 581).

More recent definitions (e.g., Nassaji’s and Goo’s) are more general in reference. They refrain from specifying meaning as the focus in interaction because the notion of meaning maintenance in recasting has been questioned. As Hauser (2005) has pointed out, “the meaning of a turn is ambiguous, indexical, and open to negotiation and renegotiation” (p. 310), only to be disambiguated through the following turns, including the turn that recasts it. In other words, the role of a recast turn may be meaning making rather than meaning maintenance. Furthermore, the greater generality of the more recent definitions also follows from the recognition of the “chameleonic-like” nature of recasts (Ellis & Sheen, 2006, p. 579). Recasts can occur with or without stress (Chaudron, 1977), in “one-signal and extended negotiations” (Braidy, 2002, p. 16); they can be corrective or noncorrective (e.g., Doughty & Varela, 1998), full or partial, single or multiple, repetitive or nonrepetitive (Ellis & Sheen, 2006), isolated or incorporated, declarative or interrogative (Lyster, 1998a), and explicit or implicit (Nassaji, 2009). In addition, recasts can be distinguished by their length (short vs. long), their linguistic focus (e.g., pronunciation, or grammar), the type of change involved (e.g., substitution, reordering), their mode (e.g., declarative or interrogative), the use of reduction (e.g., partial recasts) and the number of changes (one vs. multiple) (Sheen, 2006). Given all these variations, the best solution would be to define recasts as broadly as possible, covering only their most fundamental features (e.g., presence of targetlike reformulations, as opposed to absence of targetlike forms in “prompts”), as Ellis and Sheen (2006) have suggested. Presently, an emerging consensus seems to be that a general definition is preferred but recasts need to be operationalized and coded in view of their formal diversity.

3. The workings of recasts: expanding understanding

Cognitive interactionists posit that interaction, feedback, and output contribute to L2 acquisition. In his Interaction Hypothesis (Long, 1981, 1983, 1996), Long explained that “negotiation of meaning … facilitates acquisition because it connects input, internal learner capacities, particularly selective attention, and output in productive ways” (1996, p. 451; italics original). Gass and Mackey (2006) also suggested that “as part of this negotiation, learners receive feedback on their production, thereby potentially drawing attention to linguistic problems and leading them to notice gaps between their production and the target language” (p. 4). Similarly, Gass and Selinker (2008) suggested that interactional feedback “provides [learners] with information about the success (or, more likely, lack of success) of their utterances and gives additional opportunities to focus on production or comprehension” (p. 329). It was against this backdrop of interactionism that recasting as a type of interactional feedback strategy attracted much research attention in focus-on-form
instruction.

Recasts provide CF implicitly, do not obstruct the normal course of meaning exchange, and can be integrated naturally into communicative and task-based approaches to language teaching (Long, 2007; Long & Robinson, 1998). Recasts are also considered to be “pedagogically expeditious” and “time-saving” techniques that keep learners’ focus on meaning but at the same time allow the teacher to maintain control (Loewen & Philp, 2006, p. 537). They perform dual roles both as positive evidence (i.e., target-like forms as input) and negative evidence (i.e., correction as negative feedback in itself) (Ellis & Sheen, 2006; Mackey, 2012). The positioning of recasts side by side with erroneous forms can facilitate the noticing of the gap. That is, learners are likely to attend to the difference between their own utterance and a correct reformulation of it that ensues closely (Mackey, 2012; also see Long, 2007; Long et al., 1998). Goo and Mackey (2013, p.130) identify other advantages of the error-recast juxtaposition, noting that this sequencing (a) engenders “semantic transparency” (i.e., the comprehensibility of a recast as a reformulation of what has been said); (b) “enhances the salience of target forms and precipitates the noticing of the changes made to the learner’s original utterance,” and (c) creates opportunities for L2 learners to make a comparison of forms.

Lyster and Ranta (2013) do not directly counter the merits of recasts as summarized by Goo and Mackey. However, their objection to “the across-the-board utility of recasts” claimed by Goo and Mackey (p. 135) implies that they are skeptical about these merits. Indeed, based on their earlier work that failed to detect effects of recasts on L2 learning, Lyster and his colleagues concluded that the learning opportunities embedded in recasts are likely to be missed by L2 learners who may understand the recast turn as an integral move in communicative exchange rather than as CF (Lyster, 1998a, 1998b; Lyster & Ranta, 1997; Panova & Lyster, 2002). Their conclusion was based on their research into French immersion classrooms, where exchanges of meaning generally take precedence. Notably, in some of their recent studies (e.g., Lyster & Izquierdo, 2009; Lyster, Saito, & Sato, 2013), however, Lyster and his colleagues found recasts to be facilitative of learning, with “learners receiving recasts …benefit[ing] from repeated exposure to positive exemplars as well as from opportunities to infer negative evidence” (Lyster & Ranta, 2013, p. 175). Nonetheless, Ranta and Lyster (2007) drew on skill acquisition theory rather than the Interaction Hypothesis to explain the effectiveness of CF since the former “explicitly acknowledges a role for CF within an instructional sequence that includes language practice” (Lyster & Ranta, 2013, p. 177). This theoretical perspective, according to them, could readily account for why the effects of CF are often revealed in practice of prior knowledge.

Goo and Mackey’s (2013) theorizing about the potential of recasts to generate learning seems more enthusiastic than realistic, when research on the effectiveness of recasts has produced mixed findings. There is considerable empirical evidence suggesting that the effectiveness of recasts is contingent on the availability of learners’ attentional resources, their developmental readiness, their proficiency in the target language, learner autonomy, consistency in the focus of recasts, intensity of recasts, linguistic elements of recasts, paralinguistic elements, and explicitness of recasts (Ammar & Spada, 2006; Ellis, 2007; Han, 2002; Nassaji, 2009). Empirical evidence, especially data that trace L2 learners’
noticing and attention trajectories in both classroom and lab contexts, is especially helpful for a better understanding of the complex workings of recasting.

4. Research on uptake: disputes and consensus

Building on Lyster and Ranta’s (1997) seminal work on four primary-level French immersion classrooms, previous uptake studies measured the effectiveness of recasts by examining immediate learner responses in emergent classroom discourse. These responses were coded as uptake and repair (e.g., Lyster, 1998a, 1998b; Panova & Lyster, 2002). Uptake was defined as “a student’s utterance that immediately follows the teacher’s feedback and that constitutes a reaction in some way to the teacher’s intention to draw attention to some aspect of the student’s initial utterance” and it “reveals what the student attempts to do with the teacher’s feedback” (Lyster & Ranta, 1997, p. 49). Uptake may repair the error in question or generate an utterance that still needs repair. Based on transcripts of classroom interaction, Lyster and Ranta identified six different types of feedback used by four teachers; that is, explicit correction, recasts, clarification requests, metalinguistic feedback, elicitation, and repetition. They found recasts the most welcome option for offering feedback but the least effective in eliciting student-generated repair. Their finding about the low rate of uptake has been supported by a number of other studies (Lee, 2007; Panova & Lyster, 2002; Tsang, 2004).

Empirical evidence has accumulated in other contexts showing that recasts may trigger a fair rate of learner uptake (i.e., Braidi, 2002; Ellis, Basturkmen, & Loewen, 2001; Ohta, 2000; Oliver, 1995; Oliver & Mackey, 2003; Sheen, 2004). Using Lyster and Ranta’s taxonomy of CF moves and learner uptake, Sheen (2004) investigated four communicative instructional settings, namely French Immersion, Canada ESL, New Zealand ESL, and Korean EFL. Like Lyster and Ranta, she found recasts the most frequently used across all contexts, but she also found considerable uptake and repair in the Korean and New Zealand contexts. Sheen interpreted her findings as reflecting contextual nuances, for example, the degree of meaning-orientedness of negotiation, simplicity and explicitness of recasts, students’ orientation to form, age, proficiency, and educational background. Based on such studies, Goo and Mackey argue that recasts can lead to much higher percentages of uptake than the 18% reported by Lyster and Ranta (1997), and conclude that whether recasts lead to uptake is “a context-dependent question” (p. 137). Lyster and Ranta (2013) accept the findings from only some of the studies that reported higher uptake rates (Ellis et al., 2001; Sheen, 2004; Lyster & Mori, 2006) but regard other investigations (e.g., Braidi, 2002; Oliver, 1995; Ohta, 2000) as not being comparable to their 1997 study. They argue that Braidi’s and Oliver’s uptake figures, derived from dyadic interactions, do not undermine their earlier finding, given the difference between whole-class and pairwork and Braidi’s acknowledgement of coding ambiguity. Neither does Ohta’s (2000) finding based on four adult EFL learners’ private speech, where repetitions of recasts rather than learner-initiated repairs alone were treated as uptake. Lyster and Ranta defend the validity of their earlier finding by arguing that any statistical comparison with their
1997 study should have been made on a methodologically comparable basis. Apparently, they acknowledge the relationship between uptake rates and contextual factors, but they disagree with the conclusion that recasts can generate a substantial level of uptake.

The multitude of uptake studies notwithstanding, whether uptake is an adequate measure of learning has consistently been questioned. On the one hand, uptake may conflate superficial mimicry; on the other, learners may learn from feedback but do not respond to it externally (Mackey & Philp, 1998). Goo and Mackey (2013, p. 137) object to Lyster and Ranta’s “emphasis on uptake” and, based on a synthesis of previous arguments (Gass, 2003; Loewen & Philp, 2006; Long, 2007; Lyster & Izquierdo, 2009; Mackey & Philp, 1998; McDonough & Mackey, 2006; Ohta, 2000), draw several conclusions. First, a pretest-posttest comparison better captures the effects of recasts on L2 development or acquisition (Mackey & Philp, 1998) than the coarsely grained gauge of “the presence or absence of an overt oral response” (Ohta, 2000, p. 66). Second, primed production, defined by McDonough and Mackey (2006) as “a learner’s use of the question form provided in the recast to ask a new question” (p. 705), rather than immediate repetition (one type of uptake), predicts L2 learning. Third, the rate of successful uptake does not predict learner accuracy on tailor-made immediate and delayed posttests (Loewen & Philp, 2006). Finally, availability of opportunities to produce immediate uptake does not relate to learning, and even without uptake or repairs, recasts can facilitate L2 learning (Lyster & Izquierdo, 2009). Goo and Mackey also cite Long’s (2007) overall skepticism about the whole strand of uptake studies as support for their conclusions.

Lyster and Ranta (2013) respond that they have always maintained that “instances of uptake are not instances of learning” (p. 171; also, Lyster, 1998c, 2002, 2007; Lyster & Ranta, 1997), and that even in their 1997 work, they did leave room for further inquiry by pointing out that “claims related to language learning remain speculative and subject to further empirical investigation” (p. 57). They claim that their position regarding uptake has been misinterpreted by Goo and Mackey (2013) and Long (2007). As regards the role of uptake, however, they support the possible link suggested by Panova and Lyster (2002) between a low level of uptake following recasts and failure to process recasts as corrective. They also cite Mackey, Gass, and McDonough (2000) in support of the relevance of uptake to learner perceptions about feedback. More specifically, to argue for the usefulness of the construct, Lyster and Ranta distinguish different types of uptake (i.e., uptake with repair versus uptake without repair, and of the former, repetition versus self-repair/peer-repair) and believe that uptake in the form of a repetition contributes little to L2 acquisition (see also Panova & Lyster, 2002). They further claim that recasts can result only in learner repetitions, but prompts lead to learner-initiated repairs. The rationale for their comparison of prompts with recasts is that repetitions and learner repairs involve different processing mechanisms.

In summary, the extant uptake studies have reported both high and low levels of learner uptake, depending on the particular configurations of contextual variables, so that no definitive answer is available for the question of whether recasts are effective in leading to learner uptake. Immediate uptake may not be a valid measure of language acquisition; however, it can be a useful indicator of learner perceptions on the spot and serves as overt
evidence of learning, especially when it takes the form of self-repair or peer-repair (rather than learner repetition). Despite the controversy over the construct of uptake, the uptake studies have introduced classroom discourse analysis to feedback research, established a tentative link between interactional feedback and L2 learning, and above all, ushered in the current more direct inquiries that predominantly favor pretest-posttest designs.

5. Methodological issues in effectiveness research

5.1 An overview of empirical studies

In this section, we present a brief review of empirical studies on the (relative) effects of recasts as a prelude to our subsequent discussion on the heatedly debated methodological issues in effectiveness research on recasts. Based on the specific research designs adopted, Nassaji (2009) distinguished between experimental pretest-posttest studies and individualized test studies employing tailor-made measurements. In the former category, specific types of feedback were singled out as treatments and their respective effectiveness on preselected forms was examined. Doughty and Varela (1998), for example, investigated the effects of recasts on learning of the English past tense by examining ESL learners’ use of the target form in their lab reports prior to and after receiving implicit instruction in the form of recasts, and reported that recasts were conducive to such learning in a communicatively oriented classroom.

More common in this category were studies comparing recasts with other forms of feedback. Lyster (2004), for example, examined the effects of recasts in relation to FFI and in comparison with prompts, with grammatical gender in French as the target form. Eight classes of 179 fifth-grade students were divided into an FFI plus recast group, an FFI plus prompt group, an FFI with no feedback group and a control group receiving no FFI. Analysis of pretest, immediate-posttest, and delayed-posttest results indicated that FFI was effective, but considerably more effective when combined with prompts than with recasts or with no feedback. The difference was more manifest on the written tasks than on the oral tasks. The recast plus FFI group outperformed the control group on two after-treatment written tests and the delayed oral posttests. These findings suggested that providing prompts may be more effective than providing recasts, but recasts were obviously more effective than no feedback at all, when both types of feedback were administered along with FFI. This order of effectiveness was also reported by Ammar and Spada’s (2006) study of the third person possessive determiners in English. In an EFL setting, Yang and Lyster (2010) provided 72 Chinese learners with three types of feedback, prompts or recasts on their use of regular and irregular English past tense, or feedback on the content of their utterances only, during form-focused production activities. The pretest, immediate posttest, and delayed posttest results indicated that recasts did not appear to be as effective as prompts, because their effects were only felt on a limited number of measures and were also smaller on some of these measures.

By contrast, in a “classroom + laboratory” study, where classroom teaching of grammatical gender in French preceded the provision of either prompts or recasts in
dyadic interactions outside of the classroom, Lyster and Izquierdo (2009) recorded significant improvement in accuracy and reaction-time scores but did not detect any difference between the “prompts” and the “recasts” conditions on these measures. Pretest-posttest studies have been questioned for the intensity of feedback on the preselected form and, for this reason, for their pedagogical implications (e.g., Ellis & Sheen, 2006).

Individualized test studies investigated how learning can be promoted by recasts that treat a multitude of learner errors occurring randomly in a natural setting of interaction and measured such learning with tailor-made assessment procedures (e.g., Adams, 2007; Loewen, 2005; Nabei & Swain, 2002; Swain, 2001; Williams, 2001). Nabei and Swain, for example, designed tailor-made grammaticality judgment tests based on recast episodes they identified from class videotapes to examine how an adult female EFL learner benefited from recasts. In a similar vein, Loewen (2005) included items for an individualized test based on recast episodes identified during L2 class observation. Both studies found a positive effect of recasts in natural settings. The demerits of the individualized test approach include difficulty (and therefore lack) of pretesting (Swain, 2001; Loewen, 2005; Loewen & Philp, 2006), pseudo-errors (Braid, 2002; Loewen, 2005; Nassaji, 2009), and technical and methodological challenges. To improve on methodology, Nassaji (2009) integrated a pretest-posttest design with incidental feedback by administering a group picture description task to collect pretest data, an oral description task with teacher feedback as the individualized treatment, at-the-spot revision as an immediate posttest, and another revision task as a delayed posttest. The analyses of data showed a higher degree of immediate post-interaction correction and more pronounced effects of feedback explicitness for recasts than for elicitations.

The above review appears to support the effectiveness of recasts (Ammar & Spada, 2006; Doughty & Varela, 1998; Loewen, 2005; Nabei & Swain, 2002; Nassaji, 2009; Yang & Lyster, 2010) and recasts plus FFI (Lyster, 2004; Lyster & Izquierdo, 2009). However, with the exception of Nassaji (2009), extant studies (e.g., Ammar & Spada, 2006; Lyster, 2004; Lyster & Izquierdo, 2009; Yang & Lyster, 2010) suggested that recasts were less effective when compared with other types of feedback, for example, prompts. To argue for the effectiveness of recasts, Goo and Mackey criticize a number of comparative studies published before 2013, including those incorporating FFI, for the methodological and interpretative problems they suffered. We will turn to their criticisms next.

5.2 Opportunities for modified output
Goo and Mackey (2013) assert that in previous studies comparing prompts and recasts the recast condition was disadvantaged since participants receiving prompts were inherently pushed to produce modified output. Since “this modified output has been argued to lead learners to focus on the linguistic form in which the feedback was provided” (2013, p. 148), it would be fair to compare the effects of recasts and other types of feedback only when opportunities for modified output are blocked. They cite Goo’s (2012) study as an example, where opportunities for modified output were successfully blocked by the native speaker teacher diverting learners to the next sentence item in information gap activities (by saying “Okay, let’s move to the next item”) immediately after he provided either recasts
or metalinguistic feedback. Goo found that under this circumstance, recasts paralleled metalinguistic feedback in facilitating the acquisition of the target construction.

Lyster and Ranta put forward three counterarguments. First, they point out that since Goo and Mackey insist on recasts leading to high levels of uptake, they should expect recasts to lead to modified output as well, because modified output is analogous to learner repair (but may not be correct). Second, they note that in acknowledging that modified output opportunities are inherent in the prompt condition only, Goo and Mackey unwittingly admit the inferiority of recasts to prompts. Third, Lyster and Ranta reason that given the claimed advantages of recasts, both recasts and prompts may facilitate learning, though in different ways and, consequently, they believe that it is meaningful to compare the relative effects of these two types of feedback. According to them, “learners receiving recasts can benefit from repeated exposure to positive exemplars as well as from opportunities to infer negative evidence, whereas learners receiving prompts can benefit from repeated exposure to negative evidence as well as from opportunities to practice using the target form as they modify their output” (p. 175). Methodologically, Lyster and Ranta concur that by using the fill-in-the-blank tasks, along with teacher diversion, Goo (2012) was able to obstruct opportunities for modified output. However, they suggest that in cases where continued communication is always preferable, for example, in communicative language classrooms, such opportunities can be documented via tracking recordings (Loewen, 2004; Yang & Lyster, 2010) and should be investigated as a variable. The presence and absence of opportunities for modified output, they argue, should not be a reason for quiting comparative research.

We agree with Lyster and Ranta that the inherent inequality of types of feedback should not prevent comparative studies of the effects of recasts and prompts if the research findings are to have pedagogical relevance and practical meaningfulness. We propose that in comparative studies of recasts and prompts, opportunities for modified output can either be blocked or left open, depending on whether the focus of the study is on the effect of a recast or prompt turn per se or that of a teacher recast/prompt and learner response episode. The blocking approach may yield clearer insights into the relative effects of unalloyed recasts and prompts, but the narrow range of treatment tasks specially, if not artificially, designed to enable blocking can pose grave limitations on what can be investigated. The blocking approach is also likely to disrupt the natural flow of communication. In contrast, the non-blocking method is not subject to this problem and is more natural in a real classroom. Besides, we share Lyster and Ranta’s view that in future comparative research, the take-up of modified output opportunities should best be tracked and examined as a moderator variable.

5.3 Single versus multiple comparisons
Goo and Mackey point out that a common flaw of many comparative studies is that they operationalized prompts as diverse and multiple types of feedback (e.g., metalinguistic feedback, elicitations, repetitions, and clarification requests), provided at the teacher’s discretion (e.g., Ammar, 2008; Ammar & Spada, 2006; Yang & Lyster, 2010) in successive moves (e.g., Lyster & Izquierdo, 2009) or in a single turn (e.g., Ellis, 2007; Ellis, Loewen
& Erlam, 2006; Sheen, 2007). Such operationalization created a situation where a single treatment (i.e., recasts) was compared with multiple types of treatment (i.e., prompts). It not only disadvantaged the recast condition but also made it impossible to attribute the observed effects to a particular type of feedback.

Lyster and Ranta concede that in the studies concerned, “prompts were considered a single CF strategy, which withholds correct reformulations but has multiple manifestations” (p. 175). However, they point out that a greater variety of manifestations does not necessarily mean a greater amount of feedback. Consequently, in explaining the superior effects of prompts, they propose that “it may be their variety (but not greater frequency) that adds to their effectiveness” (p. 175), even though they do acknowledge that operationalizing prompts as a single type of feedback can enhance the rigor of comparisons and help differentiate the effects of different types of prompt.

These concessions notwithstanding, Lyster and Ranta defend the pedagogical soundness and methodological validity of the “variety” approach. They point out that recasts also “come in many shapes and sizes” (p. 175), which Mackey and Goo (2007) themselves recognize. Thus, Mackey and Goo’s “single-versus-multiple comparisons” are really “multiple-versus-multiple comparisons.” In addition, Lyster and Ranta take to task several studies reviewed approvingly by Goo and Mackey for the same operationalization issue. They point out that when Mackey and Phlip (1998) compared recasts with negotiated interaction, they operationalized the latter as clarification requests and confirmation checks and such operationalization overlapped greatly with recasts, which could function as confirmation checks. They also note that in Long, Inagaki, and Ortega’s (1998) study, which compared recasts with models, the recast condition involved more cognitive activities than the model condition since the participants receiving recasts engaged in L1 listening, translated L1 into L2, and received recasts, whereas those receiving models were only required to listen to L2 modeling before they repeated it.

The discussion above indicates that both recasts and prompts are “chameleon-like” by nature and that the operationalization of both is likely to involve multiple manifestations. From a methodological perspective, ideally, these entangled elements should be teased apart and compared individually to pinpoint their respective effects. Such more differentiated operationalization efforts can add rigor and clarity to research results and, increase the internal validity of empirical findings. From a classroom perspective, however, the “variety” approach to operationalization, by incorporating multiple feedback moves that occur naturally together, has a better chance of producing empirical findings that have ecological validity.

5.4 The varying effects of form-focused instruction
A number of comparative studies on recasts and prompts included FFI on the ground that L2 learners would need to have some knowledge about the target form to be able to respond to prompts directed at that form. As Goo and Mackey point out, however, such prior knowledge is not necessary in the case of recasts. Thus, they believe that the inclusion of FFI is likely to favor the prompt condition “because not only is the starting point likely to influence overall results but it also makes the interpretation of results regarding the
exact cause of learning somewhat difficult” (p. 152). Citing Lyster (2004), Ammar (2008), and Ammar and Spada (2006) as cases in point, Goo and Mackey suggest that FFI in such studies can serve a mediating role, rendering it difficult to attribute effects on learning to FFI, feedback, or the concerted impacts of the two. Therefore, they call for studies that do not incorporate FFI to keep all learning opportunities open so that the effects of recasts and other forms of feedback can be compared more fairly.

Lyster and Ranta do not find it problematic to include FFI in comparative studies. They believe that the complications arising from its inclusion could be resolved by providing all groups with equivalent instruction and that the provision of equal instructional activities contributes to the teasing apart of task effects from those of CF. In stark contrast to Goo and Mackey’s critique, they take Lyster (2004), Ammar (2008), and Ammar and Spada (2006) as exemplary classroom studies where the effects of prior instruction were properly controlled for by providing equivalent instructional activities such as equal practice opportunities and/or noticing and awareness tasks across groups. Lyster and Ranta agree with Li’s (2010) argument that in the case of a comparison between FFI + recasts, FFI + prompts, and FFI-only conditions, “any effect … must be due to the presence or absence of feedback” (p. 319). The opposing views held by them and Goo and Mackey have arisen from their interest in understanding “how CF fits into the bigger picture of classroom instruction,” in contrast to Goo and Mackey’s interest in “isolating the effects of CF during interaction in lab settings to control for prior knowledge” (p. 177). Given their research interest, Lyster and Ranta’s ongoing research draws on skill acquisition theory rather than the interaction hypothesis alone and focuses on how CF may help “consolidate emergent L2 knowledge and skills through practice” (p. 177).

The methodological importance of offering equivalent instructional treatments to all groups involved in comparison (e.g., recast, prompt and control) is crucial in disentangling the effects of feedback from those of instruction. Lyster and Ranta, however, have not responded to the concern over a possible interaction between instruction and types of CF (e.g., recasts vs. prompts). Such an interaction is a serious threat to the internal validity of an experimental study. For this reason, we suggest the adoption of the Solomon four-group design in further research. This experimental design allows the research to treat FFI as an independent variable and cross it with the other independent variable, types of CF, so as to determine if FFI is a moderator variable. We share Lyster and Ranta’s view that the effectiveness of CF in the consolidation of emergent L2 knowledge and skills through practice is a research topic worthy of further exploration.

5.5 Prior knowledge as a confounding variable
Goo and Mackey observe that in some previous studies (e.g., Ammar & Spada, 2006; Ellis, 2007; Loewen & Nabei, 2007), a high level of prior knowledge may have left little room for further learning of a target structure. For example, they suggest that in Ammar and Spada’s study, the lack of a significant difference on the oral production measure between the high-proficiency recast, prompt, and control groups could be attributed to the good knowledge high-proficiency learners had of the target forms. The high level of prior knowledge possessed might have left little room for further development under both
the recast and prompt conditions but benefited the control group through the provision of practice opportunities. Similarly, the lack of a difference between the high-proficiency recast and prompt groups on the passage correction task, Goo and Mackey speculate, may also be attributable to a ceiling effect as reflected in the high pretest scores on this task. Goo and Mackey also surmise that the failure to control for the effects of prior knowledge in Yang and Lyster’s (2010) study might partially explain the large pre- to posttest effect size for the control group on the oral production of the irregular past tense and the written production of regular past tense.

In Lyster and Ranta’s opinion, the purpose of providing CF should be in line with the ultimate goal of instruction, which, as they see it, is to provide “repeated opportunities [for learners] to retrieve and restructure their knowledge to become fluent and accurate users of the target language” (p. 178). Based on this understanding, they hold that studies such as Yang and Lyster (2010) which involved learners having prior knowledge of target forms were valid and meaningful because they purposively tested the facilitative roles of CF in helping learners to automatize that prior knowledge in practice, and that their findings can be rightly interpreted as such. In response to Goo and Mackey’s criticism, they counterargue that “it remains unclear how a learner with zero knowledge of a form could attempt to produce it during meaningful interaction in a way that would elicit CF from the teacher or interlocutor” (p. 178).

Goo and Mackey’s concern about prior knowledge as a confounding variable is certainly warranted if such knowledge is likely to give rise to a ceiling effect. The existence of a ceiling effect makes it impossible to tell whether the absence of significant differences between different treatment conditions is a genuine finding about their equal (in)effectiveness or a mere artifact of design and measurement problems. On the other hand, Lyster and Ranta also hold a justifiable position when they insist that a certain level of prior knowledge is a necessary precondition for participants to be able to respond to interactional feedback, especially prompts. Rigorously designed investigations into the effects of feedback on the consolidation and automatization of existing knowledge are worthwhile and have the potential to yield important pedagogical implications.

6. Directions for future research

Goo and Mackey hold that although the question of whether recasts facilitate language acquisition has been empirically settled, continued research into the effects of different forms of feedback on the acquisition of various linguistic targets is still needed, provided that such research is rigorously designed. They do not advocate comparative studies, except those with good reasons, and such reasons, they acknowledge, may include classroom considerations. They recommend that instead of comparing the relative effectiveness of recasts and prompts (or other types of CF for that matter), it will be particularly useful to explore what factors impact on their respective effectiveness. They call for more research zooming in on the effects of CF in interaction with attentional trajectories, interlocutor differences, learning environments, and new types of target.
Lyster and Ranta criticize Goo and Mackey for adopting “a reductionist research strategy” (p. 179), though they do agree that a critical examination and discussion of methodological issues is useful and serves the field well. They disagree fundamentally with Goo and Mackey on the validity and value of comparative CF studies, holding that such research should have “the interests of learners and teachers in the real world in mind” (pp. 178-179) and be both meaningful and valid. They further argue that comparative studies can help us better understand the workings of different CF techniques and help identify new variables at work. They also differ from Goo and Mackey in their support for longitudinal studies. Such studies, they argue, should not be marginalized simply for fear of the possible contamination by out-of-experiment exposure. Furthermore, they call for longitudinal CF research, especially classroom-based studies, because of their concern “with investigating SLA phenomena that are of practical significance to teaching and with conducting research in such a way that it is transparently relevant to teachers” (p. 181).

Like Lyster and Ranta, we call for a broadening of the conceptual and methodological borders of CF research if this field is to continue to yield new insights into fundamental issues of second language acquisition and to inform classroom instruction. Much needed are research endeavors that make use of new technology (e.g., recent eye-tracking techniques) to track trajectories of noticing and attention, examine the effects of recasts on learning in interaction with multiple social, cultural, and cognitive factors (e.g., working memory, social status of interlocutors, learning settings), and incorporate little-researched learning targets (e.g., L2 pragmatics), as suggested by Goo and Mackey. We additionally recommend studies investigating learner/teacher beliefs and perceptions about CF and how affective factors (e.g., anxiety, self-efficacy) may interact with learning from CF (e.g., Zhang & Rahimi, 2014). There is also room for further inquiry into the differential effects produced by finer variants of a certain feedback type (e.g., Nassaji, 2009).

Furthermore, there is a need for methodologically valid and theoretically sound comparative studies that explore the aforementioned new variables and possibilities or, as Goo and Mackey suggest, examine factors that moderate the effectiveness of different types of CF. This entails the adoption of various research designs that allow researchers to examine the interaction among multiple variables. Comparative studies in classroom settings are also needed because of their pragmatic connections to the real world and their complementary relationship to blue sky lab research. Finally, more studies in EFL settings should be conducted, not only because these settings can, as Goo and Mackey suggest, provide greater control over out-of-experiment exposure but also because findings produced in ESL contexts may not be extrapolatable to EFL ones. This is certainly an area where researchers in EFL countries can actively contribute to the knowledge base for the international SLA community and cater for local teaching needs.

7. Conclusions and implications

We have presented a critical review of the L2 literature on recasts in response to a recent debate between Goo and Mackey (2013) and Lyster and Ranta (2013). Our review makes
clear that L2 recasts research as a whole has yielded important insights into the nature of recasts, their working mechanisms, and their (relative) effects, despite the ongoing debate on the appropriate research methodology in studying recasts. Goo and Mackey’s methodological critique is valuable and does the field a great service by raising our awareness of potential methodological pitfalls and design flaws in planning and conducting further research. On the other hand, these methodological concerns should not drive out classroom-based research or studies that are closer to classroom concerns and practices. That it is challenging to design rigorous and well controlled classroom studies does not mean such studies are impossible. While the two sides of the debate disagree on many crucial issues (e.g., whether the effectiveness of recasts has been established beyond any dispute, whether it is legitimate/meaningful to conduct comparative studies of different types of CF, and whether a naturalistic approach to research based on classroom contingencies and pedagogical needs is desirable), both sides do agree on, among others, the importance of rigorous methodology in L2 recasts research, the need to have a sound theoretical grounding, and the desirability of providing multiple forms of CF in the classroom.

The shared recommendation about employing a variety of CF strategies in the classroom is particularly pertinent to an EFL context like China. To our knowledge, there has been no study following Oyster and Ranta (1997) or Sheen (2006) to investigate Chinese English teachers’ use of CF types. However, it may not be far-fetched to assume that they use a narrow range of CF techniques generally. Some of them have been observed to ignore learner errors to avoid interrupting information exchange in communicative classrooms and/or to boost learner confidence. Others have been found to overuse explicit correction (along with metalinguistic feedback) in classroom discourse, as they are used to doing when providing written feedback. Both groups of teachers need to make more strategic use of a wide range of CF strategies to cater to students’ differing needs and facilitate their learning. There are challenges to meet. To maximize the effectiveness of different types of CF in their classrooms, pre- and in-service teachers require carefully designed professional preparation that can equip them with a solid knowledge base concerning the what, why, when, where, and how of CF to draw on for appropriate feedback decisions. There are other challenges arising from resource constraints, educational structure, cultural influences, to name just a few. For example, a typical secondary English class of 50 or more students would make it infeasible for the teacher to attend to everyone’s error with recasts. Such challenges, however, are not insurmountable and provide the very impetus for researchers and teachers alike to explore alternative forms of CF in the Chinese context. Given the scope of EFL instruction in China and the surging research interest in second language acquisition in general and CF in particular, we anticipate important contributions from Chinese teachers and researchers to our growing understanding of how CF supports second or foreign language development.
References


Swain, M. (2001). Examining dialogue: Another approach to content specification and to validating


learners’ acquisition of regular and irregular past tense forms. *Studies in Second Language
Acquisition, 32,* 235-263.

Yilmaz, Y. (2012). The relative effects of explicit correction and recasts on two target structures via
two communication modes. *Language Learning, 62,* 1134-1169.


(Copyedited by Duncan Sidwell & Ding Yanren)

**About the Authors**

Chengsong YANG is a fulltime PhD candidate at The University of Auckland, New
Zealand and a lecturer at Xi’an Jiaotong University, China. His research interests lie in
second language acquisition, research methods, and sociolinguistics. Email: y.chengsong@
auckland.ac.nz

Dr Guangwei HU is Associate Professor in the English Language and Literature Academic
Group, National Institute of Education, Nanyang Technological University, Singapore. His
current research covers academic discourse, bilingualism and bilingual education, home
(bi)literacy practices and acquisition, metalinguistic awareness, and second language
acquisition. Email: guangwei.hu@nie.edu.sg

Dr. Lawrence Jun ZHANG is Associate Professor and Associate Dean, Faculty of Education,
University of Auckland, New Zealand. His interests lie in learner metacognition in
(bi)literacy, reading and writing development, representations of lexical and syntactic
knowledge in bilingual and second language acquisition, and teacher cognition and change
in language education. Email: lj.zhang@auckland.ac.nz