Studies in Second Language Acquisition, 2013, **35**, 167–184.
doi:10.1017/S027226311200071X

COUNTERPOINT PIECE: THE CASE FOR VARIETY IN CORRECTIVE FEEDBACK RESEARCH

Roy Lyster
McGill University

Leila Ranta
University of Alberta

Goo and Mackey (this issue) outline several apparent design flaws in studies that have compared the impact of different types of corrective feedback (CF). Furthermore, they argue that SLA researchers should stop comparing recasts to other types of CF because they are inherently different kinds of phenomena. Our response to their article addresses (a) the claim that the recast-learning relationship has been "settled," (b) the misleading representation of our views on uptake, (c) the characterization of the CF comparison studies as being weak and invalid, and (d) Goo and Mackey's recommendations concerning the most appropriate approach to investigating the effect of feedback on second language learning.

We welcome this opportunity to address some of the arguments and claims made by Goo and Mackey (this issue). In their article, they critically examine the methodological features of studies that have compared the efficacy of recasts to other forms of corrective feedback (CF). They argue that the current state of research supporting the benefits of recasts is broad and deep, whereas the research in support of other

Correspondence concerning this article should be addressed to Roy Lyster, Department of Integrated Studies in Education, McGill University, 3700 McTavish Street, Montreal, QC H3A 1Y2, Canada. E-mail: roy.lyster@mcgill.ca

© Cambridge University Press 2012

https://doi.org/10.1017/S027226311200071X Published online by Cambridge University Press

types of CF is weak and possibly invalid. Their criticisms of an alleged "case against recasts" center on several apparent design flaws in studies not substantiating "the across-the-board utility of recasts" (p. 135): single-versus-multiple comparisons, the confounding effect of opportunities for modified output, the provision of form-focused instruction as a feature of experimental treatments, failure to control for the prior knowledge of learners, and the impact of out-of-experiment exposure. This leads Goo and Mackey to make the rather controversial claim that SLA researchers should stop making comparisons between recasts and other forms of CF because different types of feedback are inherently different, and, therefore, apples are being compared to oranges. There are many points in the article that we dispute, but due to constraints of time and space, we limit our response to the following: (a) the claim that the recast-learning relationship has been "settled," (b) the misleading representation of our views on uptake, (c) the characterization of the CF comparison studies as being weak and invalid due to design flaws, and (d) Goo and Mackey's recommendations as to how research on feedback should be conducted in the future.

ARE RECASTS EFFECTIVE ACROSS THE BOARD?

In attributing the source of "doubt" (p. 135) regarding recast effectiveness solely to Lyster and colleagues, Goo and Mackey disregard concerns expressed previously by several applied linguists. We trace the first signs of "skepticism" (p. 136) to Corder (1967), who wrote, "simple provision of the correct form may not always be the only, or indeed the most effective form of error correction since it bars the way to the learner testing alternative hypotheses" (p. 168). Similarly, in his seminal review of the literature on error correction, Hendrickson (1978) stated that the procedure whereby teachers provide students with correct forms "is ineffective when helping students learn from their mistakes" (p. 393). On the basis of data from an early classroom observation study, Chaudron (1977) concluded that recasting—what he called "repetition with change" (but without emphasis or reduction)—was "especially weak in helping to locate the error" (p. 41).

Goo and Mackey refer to their 2007 meta-analysis as "evidence that recasts work" (p. 134). What they do not mention, however, is that the effect sizes were significantly larger for lab studies than for classroom studies. Nor do they mention having qualified their results as follows: "We cannot say they [recasts] were more or less helpful than other sorts of feedback due to the insufficient number of studies." Furthermore, they called for "more studies in order to obtain a more clear-cut picture of the effectiveness of different types of feedback" (Mackey & Goo, 2007, p. 442). It would then seem that Goo and Mackey are now

https://doi.org/10.1017/S027226311200071X Published online by Cambridge University Press

retracting their call for comparative studies, and as such we are left to wonder if this is because those that followed did not necessarily provide unwavering support for recasts. For example, Lyster and Saito's (2010) subsequent meta-analysis, which examined the effects of different types of CF specifically in second language (L2) classroom settings, found medium effect sizes for recasts and large effect sizes for prompts, which proved significantly more effective in the within-group contrasts than recasts.

Several reviews of recast studies are available that put recast effectiveness into perspective through reference to a range of linguistic, pragmatic, cognitive, and contextual constraints (e.g., Ellis & Sheen, 2006; Nicholas, Lightbown, & Spada, 2001; Sato, 2011; Sheen, 2011). For example, the positive effects for recasts appear to be reserved for learners in form-oriented settings (Nicholas et al., 2001) and for those with high literacy levels (Bigelow, delMas, Hansen, & Tarone, 2006), developmental readiness (Mackey & Philp, 1998), high working memory capacity (Mackey, Philp, Egi, Fujii, & Tatsumi, 2002), and high phonological memory, attention control, and analytic ability (Trofimovich, Ammar, & Gatbonton, 2007). Goo and Mackey acknowledge these and other constraints but then ignore them as they claim instead that research has substantiated "the across-the-board utility of recasts" (p. 135).

For the most part, research demonstrating the effectiveness of recasts has been conducted in laboratory settings where CF can be delivered intensively in consistent ways on specific linguistic targets. These laboratory studies have generally yielded positive results for recasts, but many have methodological problems—such as no control group (e.g., Ishida, 2004) or a control group receiving no CF (e.g., Han, 2002; McDonough & Mackey, 2006; Sagarra 2007)—as well as other limitations, as we will point out, akin to those that Goo and Mackey associate only with the alleged "case against recasts."

Whereas the lab studies to which Goo and Mackey appeal for support fall short of substantiating their claim for across-the-board benefits of recasts, the classroom studies to which they refer (i.e., Doughty & Varela, 1998; Goo, 2012; Loewen & Nabei, 2007) provide even less support, as we explain next. First, in Doughty and Varela's classroom study comparing CF to no CF, recasts were used solely as secondary moves in the event that the primary move, which was a prompt that repeated verbatim the learner's error, failed to elicit self-repair. Thus, the fact that Goo and Mackey refer to this study as evidence of recast effectiveness is misleading. Also worrying is their lack of criticism of this study for using multiple feedback types and not controlling for modified output, a criticism they seem to apply only to studies that appear unsupportive of recast effectiveness. Second, Goo and Mackey classify Loewen and Nabei (2007) as a classroom study, but it was not conducted in intact classrooms. Instead, it was conducted outside of the classroom, with a researcher who interacted with small groups of four learners, and so it

https://doi.org/10.1017/S027226311200071X Published online by Cambridge University Press

may be better classified as a lab study. As with other lab studies comparing different types of CF (e.g., Lyster & Izquierdo, 2009; McDonough, 2007), Loewen and Nabei found no difference across CF types.

Third, the treatment in Goo's (2012) study was a form-oriented task requiring learners to ask the teacher a question in order to complete two fill-in-the-blank exercises, each with 15 blanks. Whereas the 14 participants in the recast group repeatedly heard positive exemplars of the target form each time an error was made, the 32 participants in the metalinguistic group heard a rule, such as, "Don't use the conjunction 'that' when you ask about the subject of the subordinate clause" (p. 455), without opportunities for self-repair. We have never heard of rule repetition as a potentially effective type of CF and find it surprising that the recast group was unable to outperform the metalinguistic group. Moreover, that both groups outperformed the control group does not provide any support for recast effectiveness because the control group did not participate in the same exercises and so was unable to benefit from similar exposure to the target form. A control group in a CF comparison study needs to participate in the same treatment tasks, but without CF, as was the case in Loewen and Nabei (2007), Lyster (2004), and Yang and Lyster (2010), so that all groups have similar exposure to the target forms, with and without CF (Lyster, Saito, & Sato, 2013).

INSTANCES OF UPTAKE ARE NOT INSTANCES OF LEARNING

Goo and Mackey portray our earlier work on uptake in a particularly misleading way and fail to acknowledge that Lyster's program of research has continued to evolve since the publication of Lyster and Ranta (1997). In that seminal work, we stated upfront that, because we examined only immediate uptake, "claims related to language learning remain speculative and subject to further empirical investigation" (p. 57), and that our results needed to "be validated by means of coding sets of classroom data from other contexts" (p. 56). This call for further classroom research led to several studies that indeed revealed discrepancies in rates of uptake following recasts across instructional settings, most notably Ellis, Basturkmen, and Loewen (2001), Sheen (2004), and Lyster and Mori (2006), as well as other studies that also revealed low rates of repair: in English immersion classrooms in Korea (Lee, 2007) and English as a foreign language (EFL) classrooms in Hong Kong secondary schools (Tsang, 2004). Goo and Mackey's reference to studies of dyadic interaction (Braidi, 2002; Oliver, 1995) as counterevidence to our findings is unconvincing, especially when we consider Braidi's (2002) conclusion that, even in a laboratory setting, it was impossible to determine whether learner responses to recasts addressed form or meaning, as in, "'Yes, I recognize that that is the correct form' versus 'Yes, that is what

https://doi.org/10.1017/S027226311200071X Published online by Cambridge University Press

I meant to say,'" (p. 31) and Oliver's (2000) subsequent observation that, in contrast to dyadic interaction, "the nature of whole class interactions diminishes the opportunity for students to respond to the feedback" (p. 126). The reference to Ohta's (2000) classroom study as evidence "that L2 learners do in fact produce uptake" (Goo & Mackey, this issue, p. 137) also seems underwhelming, given that it showed only that four adult learners, who produced private speech with moderate to high frequency while wearing lapel microphones in a form-oriented foreign language classroom, tended to repeat recasts. These references—taken as evidence to suggest that the rates of uptake observed by Lyster and Ranta in French immersion classrooms were somehow wrong—completely miss the point that we now know that rates of uptake vary from context to context.

As we have always maintained, instances of uptake are not instances of learning (Lyster, 1998, 2002, 2007; Lyster & Ranta, 1997). Instead, uptake refers to a range of possible responses made by students following CF. Goo and Mackey follow Long (2007) in misrepresenting Lyster's program of research by attributing to it an "emphasis on uptake" and the "implication that uptake is a measure of learning" (p. 137). For example, they refer disapprovingly to Panova and Lyster's (2002) suggestion that "if recasts and translations are essentially corrective in purpose, there is little evidence that L2 learners in the present study processed them as such" (p. 591), which was based on Panova and Lyster's observation of minimal uptake after recasts and translations. Yet the observational data in Panova and Lyster, similar to those in Lyster and Ranta, indeed suggested that learners might not be processing the corrective function of recasts. That responses to CF may be a sign of noticing was claimed by others as well, including Mackey herself, who suggested that "uptake may be related to learners' perceptions about feedback at the time of feedback" (Mackey, Gass, & McDonough, 2000, p. 492). Concerning the value of uptake following recasts, however, Panova and Lyster noted that "uptake consisting of a repetition may not have much to contribute to L2 development, because of its redundancy in an error treatment sequence where the repair is both initiated and completed by the teacher within a single move" (p. 579).

Goo and Mackey claim in their article to have "considered the issue of uptake and its meaning" (p. 146), yet they do not explain uptake as a discourse phenomenon, and they fail to make an important distinction between (a) utterances still in need of repair or (b) utterances with repair. Utterances in need of repair include simple acknowledgements such as *yes*, hesitations, off-target responses, partial repair, and occurrences of either the same or a different error, whereas utterances with repair entail the correct reformulation of an error and thus differ from modified output, which may or may not be a correct reformulation. Learner repair can either be a repetition or self-repair, and these different types of

https://doi.org/10.1017/S027226311200071X Published online by Cambridge University Press

immediate repair are specific to one type of CF or another: Recasts and explicit correction can lead only to repetition of correct forms by students, whereas prompts can lead either to self-repair or peer repair, not to repetition. Lyster and Ranta (1997) suggested that "student-generated repairs in the error treatment sequence may be important in L2 learning" (p. 57). Thus, it was this line of research that Lyster pursued, predicting that different types CF would have differential effects on learning insofar as different types of repair (e.g., self-repair after prompts vs. repetition after recasts) entail different processing mechanisms (see Lyster, 1998, 2002, 2004).

This research has shown benefits for both recasts and prompts, with some benefits for prompts over recasts, but not to the extent predicted. For example, in the case of young immersion students, recasts were as effective as prompts in oral production measures and less effective than prompts only in written production measures (Lyster, 2004). Recasts were as effective as prompts for young English as a second language learners with high pretest scores but less effective than prompts for learners with low pretest scores (Ammar & Spada, 2006). Adult EFL students in China benefited equally from recasts and prompts in improving the accuracy of irregular past tense forms but benefited more from prompts than recasts in improving their accurate use of regular forms (Yang & Lyster, 2010). With adult Japanese learners of English, recasts were more effective than no CF for improving pronunciation of familiar lexical items but not for unfamiliar items (Saito & Lyster, 2012). In the case of Lyster and Izquierdo's (2009) study, which Goo and Mackey group with other classroom studies even though the feedback treatments were delivered individually by a researcher in the lab, all participants benefited equally from recasts and prompts. To construe these results as categorically against recasts appears reductionist, and this dichotomous argumentation has led them to gloss over important nuances that have emerged from comparative CF studies in classrooms, which we consider worthy of further pursuit.

If there were classroom researchers who were actually against recasts, they would be inclined to conceal the benefits of recasts revealed by their research and to proscribe recasts in their conclusions. In contrast, Lyster and Ranta (1997) concluded simply that "teachers might want to consider the whole range of techniques they have at their disposal rather than relying so extensively on recasts" (p. 56). Lyster and colleagues have followed suit throughout the years and have encouraged teachers to use a variety of CF types (Lyster, 2002, 2004, 2007; Lyster & Mori, 2006; Lyster & Saito, 2010; Lyster et al., 2013; Panova & Lyster, 2002). This recommendation is based on the premise that recasting what students already know is unlikely to be the most effective strategy to ensure continued L2 development and that prompting learners to draw on what they have not yet acquired is equally ineffective.

https://doi.org/10.1017/S027226311200071X Published online by Cambridge University Press

To depict Lyster as responsible for the alleged case against recasts, Goo and Mackey rely on the following quote: "The analysis of recasts as used by the four immersion teachers in the present study leads to the suggestion that, in studies investigating the effects of negative evidence on classroom SLA, recasts themselves may be red herrings" (Lyster, 1998, p. 74). Yet, by "red herrings," Lyster was specifically referring to the ways in which the then-existing classroom research had operationalized recasts as very explicit forms of CF. For example, Doughty and Varela (1998) operationalized implicit negative feedback as a repetition of the error with rising intonation and additional stress to emphasize the error, followed by a recast only if the learner failed to self-repair. Additionally, the teacher provided learners in the recast condition with written feedback on the target forms, and, as students viewed their videorecorded presentations, she paused when errors occurred and asked the class to repeat the correct forms simultaneously. Similarly, in Tomasello and Herron's (1988, 1989) studies, teachers provided recasts that were, as Long (1996) argued, "delivered in the form of explicit correction" (p. 441), wherein a teacher first wrote the incorrect form on the chalkboard, crossed it out, and then wrote the correct form while saying it out loud (i.e., the recast), before providing a brief explanation. Lyster's concern was that classroom research, claiming at that time to investigate recasts, was really investigating something other than implicit negative evidence. Goo and Mackey appear to have dissociated this concern from its original context to suggest that a case against recasts exists.

MODIFIED OUTPUT OPPORTUNITIES

Goo and Mackey argue that, in comparative studies of different types of CF, learners in prompt conditions "could be seen to be in a relatively advantageous position, all other things being equal" because they have opportunities for modified output, whereas those in recast conditions "receive feedback that rarely leads to modified output" (p. 148). First, this assertion contradicts their argument disputing "the [alleged] claim that recasts do not lead to uptake" (p. 137). Second, if recasts are as effective as they are described to be by Goo and Mackey (owing to their semantic transparency, temporal immediacy, saliency, and opportunities for comparison: see p. 130), then one may argue that learners receiving recasts are, in fact, at an advantage because they consistently receive positive evidence, whereas learners in the prompt groups do not. In any case, the purpose of comparisons is indeed to ascertain which treatment is more advantageous than others. Comparing identical rather than different types of CF would lead us back to nebulous comparisons of recasts and negotiation (see Mackey & Philp, 1998).

https://doi.org/10.1017/S027226311200071X Published online by Cambridge University Press

Prompts and recasts can be compared precisely because they are different.

On the one hand, Goo and Mackey question the role of uptake in L2 development, yet they acknowledge, on the other hand, that modified output (*learner repair* in our terms) can "lead learners to focus on the linguistic form in which the feedback was provided" (p. 148). According to this view, learner repair trumps all the advantages of recasts, and, therefore, recasts and prompts should not be compared because learners receiving prompts would have an advantage, enabling them to perform better than those receiving recasts. Thus, although Goo and Mackey extoll recasts for making "relatively few participatory demands" (p. 148) on learners, they also appear to believe that the greater participatory demands placed on learners by prompts are more beneficial. We would argue that good teaching should indeed place increased participatory demands on students, and one way to do so is by using CF types that push them to modify their output.

Goo and Mackey suggest that the ideal way to compare the effects of different types of CF would be "to carefully control production of modified output following corrective feedback—regardless of the type of feedback—such that learners produce it or are prompted to move on without producing it" (p. 148). They refer to Goo's (2012) classroom study to illustrate a context in which students were prevented from modifying their output after recasts and metalinguistic feedback alike. This was pragmatically appropriate in his study because there was no purposeful goal in the form-focused treatment task, which required university-level education students to orally produce a question to fill in a blank. Rather than fill-in-the-blank exercises, other classroom studies have tended to use more communicatively oriented tasks designed to create contexts of interaction. In such contexts, we do not recommend that students be either prevented from modifying their output or required to produce it, but we do suggest that instances of modified output be better documented than has been the case in experimental classroom studies. This would enable classroom researchers to include modified output as a variable in their interpretations (rather than terminating all comparison studies, as Goo and Mackey suggest). The extent to which learners modify their output after different types of CF in classrooms is an empirical question that can be answered by analyses of video recordings of treatment sessions (e.g., Yang & Lyster, 2010) or audio recordings of language-related episodes (e.g., Loewen, 2004).

In their lab study, Lyster and Izquierdo (2009) found recasts and prompts to be equally effective, even though modified output was allowed after prompts but not after recasts. The alleged advantage of prompts over recasts—owing to greater output opportunities—has not yet been substantiated by studies predicting such an advantage. Lyster and Izquierdo concluded that, although both types of CF proved

https://doi.org/10.1017/S027226311200071X Published online by Cambridge University Press

equally effective, the types of processing that led to their respective effectiveness were very different; that is, 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. As Lyster et al. (2013) and others have pointed out (e.g., Loewen & Philp, 2006), positive learning outcomes can be associated with self-repair following prompts but not necessarily with learner repetition following recasts, the effectiveness of which may instead be associated with specific input features such as length and prosody. These are important avenues to continue to explore in classroom contexts because, contrary to Goo and Mackey's claims, they have not yet been settled insofar as the benefits of modified output associated with prompts do not necessarily trump the benefits associated with recasts.

SINGLE-VERSUS-MULTIPLE COMPARISONS

Goo and Mackey's argument against comparing single CF moves with multiple CF moves is an important one but is again misleading when applied to classroom studies, such as those by Ammar and Spada (2006), Lyster (2004), and Yang and Lyster (2010), in which learners in the prompt groups received a greater variety of CF types but not a greater amount of CF. In these studies, prompts were considered a single CF strategy, which withholds correct reformulations but has multiple manifestations (e.g., clarification requests, repetition of learner error, elicitation, metalinguistic clues). Because prompts include more than one way of providing negative evidence while withholding positive evidence, we agree that it may be their variety (but not greater frequency) that adds to their effectiveness. If it were feasible and desirable for a teacher to provide only one type of prompt repeatedly, then this refinement may add more rigor to the comparisons and would help tease apart the effects of prompts considered implicit (e.g., clarification requests) from those considered more explicit. We hasten to add, however, that we consider Goo's (2012) operationalization of prompts as a single rule explanation repeated after every target error to be questionable in terms of pedagogical appropriateness. Variety in the use of prompts may be considered more pedagogically sound and also methodologically valid in comparative studies when we consider that recasts, too, come in many shapes and sizes (e.g., Loewen & Philp, 2006; Sheen, 2006). Mackey and Goo (2007) themselves acknowledged that recasts are "increasingly recognized as being elastic in nature," although they considered them, as many researchers do, "as a monolithic construct"

https://doi.org/10.1017/S027226311200071X Published online by Cambridge University Press

(p. 413). Of interest to pursue in this regard is that variety of CF—rather than quantity—may play a major role in its effectiveness (see Lyster et al., 2013).

Goo and Mackey's criticism of single-versus-multiple comparisons seems only to apply, however, if the comparison did not come out in favor of recasts. For example, they refer to Mackey and Philp's (1998) study to support recast effectiveness, but this study compared recasts with negotiated interaction, which contains both clarification requests and confirmation checks; negotiated interaction thus combined input-providing and output-pushing types of CF and also overlapped with recasts, many of which, in their own data, proved to "function as confirmation checks" and thus to occur as "part of negotiation sequences" (Mackey & Philp, 1998, p. 342; see also Loewen & Philp, 2006; Lyster, 1998; Sheen & Ellis, 2011).

Similar support for recast effectiveness is attributed to the Long, Inagaki, and Ortega (1998) lab study that compared recasts with models. Learners in both the recast and model treatments were equipped with headphones as they played a communication game in dyads with a researcher. In the recast condition, they were told in their first language (L1) to perform a specific task. As they performed the task, they had to translate the directive to describe their action in their L2 to the researcher, who then recast their utterance. In the model condition, learners heard a prerecorded utterance in the L2 and then repeated the utterance for the researcher, who responded by performing the action. Comparing the effects on learners of receiving either (a) a recast after translating from the L1 to the L2 or (b) no CF after repeating a model does not seem to us to be a particularly robust and valid way of assessing recast effectiveness, given the additional processing involved in the recast condition (i.e., listen to the L1 utterance, translate into the L2, and receive recast) versus the model condition (i.e., listen to the L2 utterance and repeat it). Nonetheless, this study, too, was exempt from Goo and Mackey's criticism of design issues that affect comparability.

FORM-FOCUSED INSTRUCTION

Although theoretical justifications abound for isolating the effects of CF from instruction, Goo and Mackey's suggestion that CF must be provided in the absence of instruction even in classroom research shows a lack of understanding of instructional practices and also research designs. Without equivalent instructional activities across all groups, including a control group, studies cannot tease apart CF effects from task effects, as in Goo's (2012) study comparing two CF groups, who completed an instructional task while receiving either recasts or rules, with a control group, who did not engage in the same task. Learners in

https://doi.org/10.1017/S027226311200071X Published online by Cambridge University Press

the control group missed equivalent exposure to positive exemplars of the target form, which, Goo acknowledged, was a "rare" and "limited" form unless targeted by instruction (p. 453).

In classroom settings, we believe that CF is best provided strategically with form-focused instruction and that there are ways to experimentally tease apart the effects of CF from those of form-focused instruction. For example, in the classroom studies by Ammar and Spada (2006), Lyster (2004), and Yang and Lyster (2010), the CF and no-CF groups alike were all exposed to the same instructional activities, which included focused production tasks that provided practice opportunities in all groups and contexts for CF provision in the treatment groups. To integrate these treatment tasks into immersion students' regular curriculum, Lyster's (2004) study also included various noticing and awareness tasks. Engaging all groups in these tasks served to level the playing field for comparisons between CF and no-CF groups and also to ensure that those receiving CF were at a similar level in their knowledge of the target feature. As Li (2010) argued, because the no-CF group, which engaged in the same instructional activities, served as a control group, any differences detected among the recast, prompt, and no-CF groups "must be due to the presence or absence of feedback" (p. 319).

We believe that the design of these CF comparison studies is a strength because we are interested in how CF fits into the bigger picture of classroom instruction. In contrast, Goo and Mackey are less interested in the role of CF in instructional practices and more interested in isolating the effects of CF during interaction in lab settings to control for prior knowledge. Advocating a range of research approaches to understanding the complex role and variable effects of different types of CF would seem more fruitful than advocating one program of research associated with one specific theoretical orientation. In our more recent writing on CF (Ranta & Lyster, 2007), we have drawn more from skill acquisition theory than from the interaction hypothesis to explain CF effectiveness because skill acquisition theory explicitly acknowledges a role for CF within an instructional sequence that includes language practice (see also Lyster & Sato, in press). The purpose of CF in educational settings is for learners not only to notice target exemplars in the input but also to consolidate emergent L2 knowledge and skills through practice.

PRIOR KNOWLEDGE

Goo and Mackey criticize a set of classroom studies for not controlling for prior knowledge. In so doing, they imply that research in support of the positive effects for recasts has consistently controlled for prior knowledge, which is not the case. They appear to have been influenced

https://doi.org/10.1017/S027226311200071X Published online by Cambridge University Press

by Long (2007), who claimed that, with regard to the purpose of CF, "acquisition of new knowledge is the major goal, not 'automatizing' the retrieval of existing knowledge" (p. 102). In our view, the ultimate goal of instruction is not to continuously present only new knowledge to students without sufficiently providing subsequent opportunities for assimilation and consolidation of that knowledge; rather, students need repeated opportunities to retrieve and restructure their knowledge to become fluent and accurate users of the target language. Goo and Mackey miss this point as they criticize the Yang and Lyster (2010) study for its interactional activities designed to elicit target features because they "might have evoked learners' prior knowledge of the targets and have provided practice opportunities for learners in the control group even without corrective feedback" (p. 155). The study was with Chinese EFL learners at the postsecondary level (i.e., English majors), who, of course, had had plenty of prior knowledge of past tense forms—as Yang and Lyster acknowledged—but who might have lacked communicative practice in using them. That learners had prior knowledge of the target forms does nothing to detract from the results showing (a) positive effects for learners engaging in practice activities without CF, (b) more positive results for those receiving recasts, and (c) even better results for those receiving prompts (depending on the linguistic structure).

Goo and Mackey's recommendation that CF studies should "examine L2 targets to which learners have never been exposed" but "without providing instruction" (p. 153) is inadvisable not only from an educational perspective but also from a research perspective. We see little point in providing CF on forms to which learners have not yet been exposed, mainly because 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 (see Lyster et al., 2013).

HOW SHOULD CF RESEARCH BE CONDUCTED?

Apples and Oranges?

Our last point of discussion relates to Goo and Mackey's pronouncements about how CF research in the future should be carried out. Their main argument is that the empirical comparison of recasts and prompts is essentially a comparison between apples and oranges. In their view, because recasts and prompts differ in the degree to which they invite a modified response and serve different processing functions, any comparison between them is inappropriate. We strongly disagree with Goo and Mackey's assertion that CF comparison research is misdirected. In our view, such research has the interests of learners and teachers in

https://doi.org/10.1017/S027226311200071X Published online by Cambridge University Press

the real world in mind. We further dispute that such comparisons are invalid. As is true in so many aspects of life, through direct comparison, we gain a better grasp of the theoretical underpinnings of how recasts and prompts operate. Although Goo and Mackey criticize the studies by Yang and Lyster (2010) and Ammar (2008; Ammar & Spada, 2006) for failing to control for learner proficiency, it can just as easily be asserted that it is comparative studies like these that lead us to an understanding that recasts and prompts may be differentially effective depending on learners' prior knowledge. Additional studies can be designed to pursue this relationship explicitly.

Lab Research Only?

Goo and Mackey have adopted a reductionist research strategy derived from the natural sciences whereby phenomena are reduced to their constituent parts and thus allow the researcher "to find points of entry into otherwise impenetrably complex systems" (Wilson, 1998, p. 59). They recommend further investigation of how CF interacts with such factors as (a) noticing and attention, (b) individual differences in cognitive capacities, and (c) the type of target. We agree that such a program of research is likely to offer some insights into the nature of human interaction. Furthermore, critical examination of research design issues, as Goo and Mackey have attempted, serves to sensitize novice and experienced researchers to important considerations when planning studies and interpreting findings. We also agree that it is likely that research carried out in foreign language settings in which the target language is not in wide use in the school and community will be able to minimize the out-of-experiment exposure threat to internal validity.

We part company with Goo and Mackey, however, when their overriding concern with internal validity leads them to de-emphasize longitudinal research—because of the inherent threat of out-of-experiment exposure—and to recommend lab research over classroom research. Devaluing longitudinal research is the wrong message to send out to novice researchers due to the fact that a long-term view of learning is essential for addressing many important SLA issues (Ortega & Iberri-Shea, 2005). Goo and Mackey argue that because (a) internal validity trumps external or ecological validity and (b) real-world settings do not allow adequate control over variables, lab research is the only option. We believe (along with many others, including Mackey & Gass, 2005) that both lab- and classroom-based SLA research are needed and that the two complement each other. When findings converge, we can be relatively certain of having understood the phenomenon at hand. For example, both lab and classroom research have demonstrated

https://doi.org/10.1017/S027226311200071X Published online by Cambridge University Press

positive effects for form-focused instruction (Norris & Ortega, 2000; Spada, 2011). Unfortunately, as we have seen, CF in lab settings and CF in classrooms do not always operate similarly. Goo and Mackey conclude that research design flaws in classroom research are to blame. However, it could be the case that performing interactive tasks in the lab and performing them in one's own language classroom are psychologically distinct experiences for the learner. Although Gass, Mackey, and Ross-Feldman (2005) provide empirical evidence that one population of university learners behaved in the same way in a lab setting as in their foreign language classroom, other studies have shown differences in learners' interactive behavior (Eckerth, 2009; Foster, 1998). Unlike Goo and Mackey, we would argue that if lab and classroom research findings do not converge, more classroom research is needed—not less.

CF RESEARCH AND L2 TEACHING

To push the apples-and-oranges metaphor a bit further, we would argue that teachers are concerned with the "nutritional needs of learners" (Schachter, 1983) and thus need to know when apples are a better choice over oranges. According to Schachter, one nutritional need of learners is negative input (i.e., evidence). For language teachers, the framing of this issue is in terms of how they should respond to learners' errors in speaking. Surveys of teachers in different contexts have shown that teachers feel that CF is important (Schulz, 2001) and that teachers prefer to recast learners' errors (Bell, 2005). This is possibly due to the influence of teacher-education literature and courses in which the potential negative affective impact of explicit forms of CF is often stressed (e.g., Harmer, 2007). In our own pre- and in-service teacher-education courses and professional development workshops, we emphasize the need for the use of a variety of CF moves (Lyster, 2007; Ranta, in press), as do Ellis (2009) and Goo and Mackey (this issue, p. 132). Importantly, our advice to teachers is not based on an opinion or belief but, rather, on empirical evidence from CF comparison studies carried out in classrooms whose overall characteristics we are able to describe. There is reason to believe that teachers are likely to be influenced by evidence that comes from teaching and learning settings comparable to their own. Hence it is necessary for classroom research to be conducted in a wide variety of settings rather than considering that the issue has been settled.

CONCLUSION

In our response to Goo and Mackey's article, we have offered counter-arguments to their critical comments, which we hope will provide food

https://doi.org/10.1017/S027226311200071X Published online by Cambridge University Press

for thought for all readers of *SSLA*. From this exercise, it is clear to us that the CF research literature is so broad and ranging that it is inadvisable to make generalizations about the benefits of any particular type of CF. Although we clearly do not view the goals of CF research nor the findings from the existing body of literature on CF in the same way as Goo and Mackey, we believe that this critical debate serves an important awareness-raising function. As so many scholars have noted, both lab and classroom research are needed to further our understanding of SLA. We wish Goo and Mackey well in the pursuit of their proposed research agenda that seeks to map out the workings of CF through an analytic research strategy that prioritizes internal validity. For our part, as applied SLA researchers (Ellis, 2011; Kramsch, 2000), we are concerned 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.

Received 18 August 2012

REFERENCES

Ammar, A. (2008). Prompts and recasts: Differential effects on second language morphosyntax. *Language Teaching Research*, *12*, 183–210.
Ammar, A., & Spada, N. (2006). One size fits all? Recasts, prompts and L2 learning. *Studies in Second Language Acquisition*, *28*, 543–574.
Bell, T. (2005). Behaviors and attitudes of effective foreign language teachers: Results of a questionnaire study. *Foreign Language Annals*, *38*, 259–270.
Bigelow, M., delMas, R., Hansen, K., & Tarone, E. (2006). Literacy and the processing of oral recasts in SLA. *TESOL Quarterly*, *40*, 665–689.
Braidi, S. M. (2002). Reexamining the role of recasts in native-speaker/nonnative-speaker interactions. *Language Learning*, *52*, 1–42.
Chaudron, C. (1977). A descriptive model of discourse in the corrective treatment of learners' errors. *Language Learning*, *27*, 29–46.
Corder, S. P. (1967). The significance of learners' errors. *International Review of Applied Linguistics*, *4*, 161–170.
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.
Eckerth, J. (2009). Negotiated interaction in the L2 classroom. *Language Teaching*, *42*, 109–130.
Ellis, R. (2009). Corrective feedback and teacher development. *L2 Journal*, *1*, 3–18. Retrieved from: http://escholarship.org/uc/item/2504d6w3
Ellis, R. (2011). Second language acquisition, teacher education and language pedagogy. *Language Teaching*, *43*, 182–201.
Ellis, R., Basturkmen, H., & Loewen, S. (2001). Learner uptake in communicative ESL lessons. *Language Learning*, *51*, 281–318.
Ellis, R., & Sheen, Y. (2006). Reexamining the role of recasts in second language acquisition. *Studies in Second Language Acquisition*, *28*, 575–600.
Foster, P. (1998). A classroom perspective on the negotiation of meaning. *Applied Linguistics*, *19*, 1–23.
Gass, S., Mackey, A., & Ross-Feldman, L. (2005). Task-based interactions in classroom and laboratory settings. *Language Learning*, *55*, 575–611.
Goo, J. (2012). Corrective feedback and working memory capacity in interaction-driven L2 learning. *Studies in Second Language Acquisition*, *34*, 445–474.

https://doi.org/10.1017/S027226311200071X Published online by Cambridge University Press

Han, Z. (2002). A study of the impact of recasts on tense consistency in L2 output. *TESOL Quarterly*, *36*, 543–572.
Harmer, J. (2007). *The practice of English language teaching* (4th ed.). London: Longman.
Hendrickson, J. (1978). Error correction in foreign language teaching: Recent theory, research, and practice. *Modern Language Journal*, *62*, 387–398.
Ishida, M. (2004). Effects of recasts on the acquisition of the aspectual form of *-te i-(ru)* by learners of Japanese as a foreign language. *Language Learning*, *54*, 311–394.
Kramsch, C. (2000). Second language acquisition, applied linguistics, and the teaching of foreign languages. *Modern Language Journal*, *84*, 311–326.
Lee, J. (2007). Corrective feedback and learner uptake in English immersion classrooms at the primary level in Korea. *English Teaching*, *62*, 311–334.
Li, S. (2010). The effectiveness of corrective feedback in SLA: A meta-analysis. *Language Learning*, *60*, 309–365.
Loewen, S. (2004). Uptake in incidental focus on form in meaning-focused ESL lessons. *Studies in Second Language Acquisition*, *54*, 153–183.
Loewen, S., & Nabei, T. (2007). Measuring the effects of oral corrective feedback on L2 knowledge. In A. Mackey (Ed.), *Conversational interaction in second language acquisition: A collection of empirical studies* (pp. 361–377). Oxford: Oxford University Press.
Loewen, S., & Philp, J. (2006). Recasts in the adult English L2 classroom: Characteristics, explicitness, and effectiveness. *Modern Language Journal*, *90*, 536–556.
Long, M. (1996). The role of the linguistic environment in second language acquisition. In W. Ritchie & T. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 413–468). San Diego, CA: Academic Press.
Long, M. (2007). *Problems in SLA*. Mahwah, NJ: Erlbaum.
Long, M., Inagaki, S., & Ortega, L. (1998). The role of implicit negative evidence in SLA: Models and recasts in Japanese and Spanish. *Modern Language Journal*, *82*, 357–371.
Lyster, R. (1998). The ambiguity of recasts and repetition in L2 classroom discourse. *Studies in Second Language Acquisition*, *20*, 51–81.
Lyster, R. (2002). Negotiation in immersion teacher-student interaction. *International Journal of Educational Research*, *37*, 237–253.
Lyster, R. (2004). Differential effects of prompts and recasts in form-focused instruction. *Studies in Second Language Acquisition*, *26*, 399–432.
Lyster, R. (2007). *Learning and teaching languages through content: A counterbalanced approach*. Amsterdam: Benjamins.
Lyster, R., & Izquierdo, J. (2009). Prompts versus recasts in dyadic interaction. *Language Learning*, *59*, 453–498.
Lyster, R., & Mori, H. (2006). Interactional feedback and instructional counterbalance. *Studies in Second Language Acquisition*, *28*, 269–300.
Lyster, R., & Ranta, L. (1997). Corrective feedback and learner uptake: 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.
Lyster, R., Saito, K., & Sato, M. (2013). State-of-the-art article: Oral corrective feedback in second language classrooms. *Language Teaching*, *46*, 1–40.
Lyster, R., & Sato, M. (in press). Skill acquisition theory and the role of practice in L2 development. In M. G. Mayo, J. Gutierrez-Mangado, & M. M. Adrián (Eds.), *Multiple perspectives on second language acquisition*. Amsterdam: Benjamins.
Mackey, A., & Gass, S. (2005). *Second language research: Methodology and design*. Mahwah, NJ: Erlbaum.
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. 407–452). Oxford: Oxford University Press.
Mackey, A., & Philp, J. (1998). Conversational interaction and second language development: Recasts, responses, and red herrings? *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. 181–209). Amsterdam: Benjamins.

https://doi.org/10.1017/S027226311200071X Published online by Cambridge University Press

McDonough, K. (2007). Interactional feedback and the emergence of simple past activity verbs in L2 English. In A. Mackey (Ed.), *Conversational interaction in second language acquisition: A collection of empirical studies* (pp. 323–338). Oxford: Oxford University Press.
McDonough, K., & Mackey, A. (2006). Responses to recasts: Repetitions, primed production, and linguistic development. *Language Learning*, *56*, 693–720.
Nicholas, H., Lightbown, P., & Spada, N. (2001). Recasts as feedback to language learners. *Language Learning*, *51*, 719–758.
Norris, J., & Ortega, L. (2000). Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language Learning*, *50*, 417–528.
Ohta, A. S. (2000). Rethinking recasts: A learner-centered examination of corrective feedback in the Japanese language classroom. In J. K. Hall & L. S. Verplaeste (Eds.), *The construction of second and foreign language learning through classroom instruction* (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. (2000). Age differences in negotiation and feedback in classroom and pairwork. *Language Learning*, *50*, 119–151.
Ortega, L., & Iberri-Shea, G. (2005). Longitudinal research in second language acquisition: Recent trends and future directions. *Annual Review of Applied Linguistics*, *25*, 26–45.
Panova, I., & Lyster, R. (2002). Patterns of corrective feedback and uptake in an adult ESL classroom. *TESOL Quarterly*, *36*, 573–595.
Ranta, L. (in press). Feedback on spoken errors. *Moving professional learning to classroom practice: An instructor handbook*. Toronto: TESL Ontario.
Ranta, L., & Lyster, R. (2007). A cognitive approach to improving immersion students' oral language abilities: The Awareness-Practice-Feedback sequence. In R. DeKeyser (Ed.), *Practice in a second language: Perspectives from applied linguistics and cognitive psychology* (pp. 141–160). New York: Cambridge 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 collection of empirical studies* (pp. 229–248). Oxford: Oxford University Press.
Saito, K., & Lyster, R. (2012). Effects of form-focused instruction and corrective feedback on L2 pronunciation development of /ɹ/ by Japanese learners of English. *Language Learning*, *62*, 595–633.
Sato, M. (2011). Constitution of form-orientation: Contributions of context and explicit knowledge to learning from recasts. *Canadian Journal of Applied Linguistics*, *14*, 1–28.
Schachter, J. (1983). Nutritional needs of language learners. In M. Clarke & J. Handscombe (Eds.), *On TESOL '82: Pacific perspectives on language learning and teaching* (pp. 175–189). Washington, DC: TESOL.
Schulz, R. (2001). Cultural differences in student and teacher perceptions concerning the role of grammar instruction and corrective feedback: USA-Colombia. *Modern Language Journal*, *85*, 244–258.
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*, *8*, 361–392.
Sheen, Y. (2011). *Corrective feedback, individual differences and second language learning*. New York: Springer Verlag.
Sheen, Y., & Ellis, R. (2011). Corrective feedback in language teaching. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (Vol. *2*, pp. 593–610). London: Routledge.
Spada, N. (2011). Beyond form-focused instruction: Reflections on past, present and future research. *Language Teaching*, *44*, 225–236.
Tomasello, M., & Herron, C. (1988). Down the garden path: Inducing and correcting overgeneralization errors in the foreign language classroom. *Applied Psycholinguistics*, *9*, 237–246.
Tomasello, M., & Herron, C. (1989). Feedback for language transfer errors: The garden path technique. *Studies in Second Language Acquisition*, *11*, 385–395.

https://doi.org/10.1017/S027226311200071X Published online by Cambridge University Press

Trofimovich, P., Ammar, A., & Gatbonton, E. (2007). How effective are recasts? The role of attention, memory, and analytic ability. In A. Mackey (Ed.), *Conversational interaction in second language acquisition: A collection of empirical studies* (pp. 171–195). Oxford: Oxford University Press.

Tsang, W. (2004). Feedback and uptake in teacher-student interaction: An analysis of 18 English lessons in Hong Kong secondary classrooms. *Regional Language Centre Journal*, *35*, 187–209.

Wilson, E. O. (1998). *Consilience: The unity of knowledge*. New York: Vintage Books.

Yang, Y., & Lyster, R. (2010). Effects of form-focused practice and feedback on Chinese EFL learners' acquisition of regular and irregular past tense forms. *Studies in Second Language Acquisition*, *32*, 235–263.

https://doi.org/10.1017/S027226311200071X Published online by Cambridge University Press