Studies in Second Language Acquisition, 2010, **32**, 265–302.
doi:10.1017/S0272263109990520

ORAL FEEDBACK IN CLASSROOM SLA

A Meta-Analysis

Roy Lyster and Kazuya Saito
McGill University

To investigate the pedagogical effectiveness of oral corrective feedback (CF) on target language development, we conducted a meta-analysis that focused exclusively on 15 classroom-based studies (N = 827). The analysis was designed to investigate whether CF was effective in classroom settings and, if so, whether its effectiveness varied according to (a) types of CF, (b) types and timing of outcome measures, (c) instructional setting (second vs. foreign language classroom), (d) treatment length, and (e) learners' age. Results revealed that CF had significant and durable effects on target language development. The effects were larger for prompts than recasts and most apparent in measures that elicit free constructed responses. Whereas instructional setting was not identified as a contributing factor to CF effectiveness, effects of long treatments were larger than those of short-to-medium treatments but not distinguishable from those of brief treatments. A simple regression analysis revealed effects for age, with younger learners benefiting from CF more than older learners.

This study was partially funded by the Social Sciences and Humanities Research Council of Canada (410-2006-2212) and was presented at the Second Language Research Forum in Lansing, MI, in October 2009. We are grateful to Leila Ranta, Younghee Sheen, Masatoshi Sato, and two anonymous *SSLA* reviewers for helpful comments on an earlier version.

Address correspondence to: Roy Lyster, Department of Integrated Studies in Education, McGill University, 3700 McTavish St., Montreal, QC H3A 1Y2, Canada; e-mail: roy.lyster@mcgill.ca; or Kazuya Saito, Department of Integrated Studies in Education, McGill University, 3700 McTavish St., Montreal, QC H3A 1Y2, Canada; e-mail: kazuya.saito@mail.mcgill.ca.

From theoretical perspectives, the effectiveness of corrective feedback (CF) on learners' interlanguage development has been the topic of much discussion in SLA research. The effectiveness of CF lends support to SLA theorists who emphasize the importance of negative as well as positive evidence in second language (L2) development (Gass, 1997; Long, 1996, 2007) as opposed to those who argue that positive evidence alone is sufficient (Krashen, 1981) and that negative evidence in the form of CF can even be detrimental to interlanguage development (Truscott, 1999). Theoretically motivated interactional studies have typically examined conversational moves triggered by communication breakdown and message incomprehensibility (i.e., negotiation for meaning). However, the application of findings of dyadic conversation to L2 classrooms can only be indirect at best. As Lyster and Mori (2006) argued, "teacher-student interaction has a clearly pedagogical focus that relates not only to meaning but also to formal accuracy, quality of expression, and literacy development" (p. 278; i.e., negotiation of form). To further inform theoretical discussions of oral CF and especially its educational value in L2 classrooms, this article presents a meta-analysis of 15 quasi-experimental CF studies (N = 827) that involve teacher-student interaction in classroom settings.

CORRECTIVE FEEDBACK

An increasing number of SLA studies show that CF plays a role in L2 learners' interlanguage development. Two recent meta-analysis studies provide helpful findings for future studies in this vein: Mackey and Goo (2007) conducted a meta-analysis of 28 interaction studies (including 20 oral CF studies), and Russell and Spada (2006) performed a meta-analysis of 15 CF studies (including 10 oral CF studies). Russell and Spada found that CF is facilitative of L2 development; they identified a very large effect size of 1.16.[1] Similarly, Mackey and Goo discovered that providing CF in L2 interaction has a medium effect size of 0.71 in immediate posttests and a large effect size of 1.09 in delayed posttests. These results support the consensus that focus on form through CF is beneficial. Russell and Spada concluded, however, that "the wide range of variables examined in CF research is spread rather thin; more work is needed to consolidate efforts and focus on those CF variables that appear to be particularly fruitful for future investigation" (p. 156).

Research Setting

One of the key variables known to differentially affect CF effectiveness is whether the research is conducted in a laboratory or a classroom

setting. Previous review articles (Ellis & Sheen, 2006; Nicholas, Lightbown, & Spada, 2001; Spada, 1997; Spada & Lightbown, 2009) pointed out that the effectiveness of CF and subsequent learning outcomes are notably different in laboratory and classroom settings, and the meta-analysis of interaction studies by Mackey and Goo (2007) confirmed that empirical data from classroom and laboratory settings generate essentially different results. They reported medium effect size values for classroom studies (d = 0.57 for immediate posttests and d = 0.76 for long delayed posttests) and much larger effect sizes for laboratory studies (d = 0.96 for immediate posttests and d = 1.20 for long delayed posttests). Learners in laboratory settings may be more sensitive to CF because variables such as intensity and consistency are well controlled for; in the case of classroom settings, teachers may have difficulty delivering CF following specific linguistic targets in a consistent manner (Ellis & Sheen; Nicholas et al.). To reveal CF patterns with potential pedagogical implications, we chose to focus exclusively on classroom studies in the present meta-analysis. As Spada and Lightbown recently remarked, "classroom-based studies are most likely to lead to a better understanding about the kind of interaction that occurs in classrooms where the teacher is the only proficient speaker and interacts with a large number of learners" (p. 159).

CF Types

Much variation in the operationalization of supposedly similar constructs across CF studies has been noted in previous meta-analyses, one of the most salient problems being the definitional fuzziness of different types of CF. Russell and Spada (2006) drew no conclusions about the relative effectiveness of CF types, acknowledging that, despite the increasing number of observational and experimental studies on the efficacy of various types of CF, sufficient data to arrive at such conclusions are lacking. Similarly, Mackey and Goo (2007) emphasized "the need for greater theoretical specificity or practical motivations in making claims about the superiority of one feedback type over another" (p. 440).

Based on the growing body of research on CF in classroom settings, a comparison of three different types of CF is included in this meta-analysis: recasts, explicit correction, and prompts. In that way, we consider feedback types not only in terms of implicitness and explicitness but also in terms of Ranta and Lyster's (2007) distinction between reformulation, which "includes recasts and explicit correction because both these moves supply learners with target reformulations of their non-target output" (p. 152), and prompts, which "include a variety of signals, other

than alternative reformulations, that push learners to self-repair" (p. 152). Lyster and Ranta (1997) defined recasts as "the teacher's reformulation of all or part of a student's utterance, minus the error" (p. 46). Explicit correction also provides the correct form but, unlike recasts, "clearly indicates that what the student had said was incorrect" (p. 46). Prompts include elicitation, metalinguistic clues, clarification requests, and repetition. Although prompts include a range of CF types, they have one important feature in common that differentiates them from reformulation moves: They withhold correct forms and instead provide clues to prompt students to retrieve these correct forms from their existing knowledge. In terms of linguistic evidence, therefore, explicit correction provides both negative and positive evidence, recasts provide positive evidence and possibly also negative evidence, and prompts provide only negative evidence.

Because prompts can include both implicit and explicit CF moves, Ellis (2006) suggested that differentiating implicit and explicit strategies in studies of CF effectiveness would more aptly reflect "current interest in the relative contributions of implicit and explicit feedback to acquisition" (p. 29). However, limiting comparisons of CF effectiveness in L2 classrooms to only implicit versus explicit distinctions in this way (a) assumes the existence of nonoverlapping definitions of CF types that might permit categorical comparisons of implicit versus explicit interventions and (b) overestimates both the theoretical value and feasibility of comparing implicit and explicit CF types in classroom settings. Researchers have often differentiated CF types in terms of explicitness, but this has proven problematic. Recasts, for example, are often considered categorically implicit (Long, 1996; Long & Robinson, 1998), yet research shows that, depending on discourse context, instructional setting, and learner orientation as well as formal characteristics such as linguistic targets, length, and number of changes, recasts can also be quite explicit (Ellis & Sheen, 2006; Nicholas et al., 2001; Sheen, 2004, 2006). Similarly, CF considered explicit also comes in many shapes and sizes. Some explicit feedback techniques include provision of the correct form (e.g., Lyster & Ranta, 1997; Sheen, 2007), whereas other techniques withhold the correct form and either provide metalinguistic information (e.g., Ellis, 2007) or simply a metalinguistic clue (i.e., not an explanation) that indicates that the learner's utterance is ill-formed (e.g., Loewen & Nabei, 2007).

Prompts range from implicit to explicit but are distinguishable from recasts and explicit correction in terms of what Ortega (2009) called *demand*—that is, "the degree of conversational urgency exerted upon interlocutors to react to the negative feedback" (p. 75). Prompts are not necessarily explicit in terms of the linguistic information they provide but might be considered explicit in terms of their illocutionary force. In other words, by prompting, a teacher provides cues for learners to draw

on their own resources to self-repair, whereas by providing explicit correction or recasting, a teacher both initiates and completes a repair within a single move. Prompts thus fit well with instructional discourse, as they resemble the clueing procedure identified by McHoul (1990) in his study of feedback in subject-matter classrooms. *Clueing* is a "withholding phenomenon [...] different from that encountered in conversation" (p. 355), whereby teachers show students where their talk is in need of correction and not how the correction should be made. McHoul found that teachers tended to provide correct responses only as a last resort when the clueing failed to elicit self-repair. Because the main purpose of our meta-analysis is to investigate the effectiveness of CF in classroom settings, the comparison of prompts that withhold correct forms with other CF types that provide correct forms has considerable ecological validity because teachers are able to make online decisions about whether to withhold or to provide correct forms in response to students' errors (i.e., prompts vs. reformulation), but they may not be able to make online decisions about whether to deliver CF either explicitly or implicitly. This is especially true given the wide range of variables that interact to determine degrees of explicitness from a learner's perspective. That prompts comprise four specific types of CF also has ecological validity in the sense that it remains more feasible for teachers to select from a range of prompts than to consistently use only one type (e.g., only clarification requests or only repetition of error).

Recasts, explicit correction, and prompts are compared here because their definitions are sufficiently categorical to allow for relatively clear comparisons. Other classifications of CF types, notably those that distinguish between recasts and negotiation in many studies of interaction (e.g., Iwashita, 2003; Mackey, 2006; Mackey & Philp, 1998; Oliver, 2000), are less categorical for two reasons. First, negotiation for meaning conflates CF types that are input-providing and output-pushing (to use terminology from Ellis, 2006) such as confirmation checks and clarification requests, respectively, thus precluding analysis of the effects of different types of processing. Second, because negotiation for meaning includes confirmation checks, it overlaps with recasts, many of which "are often part of negotiation sequences and function as confirmation checks" (Loewen & Philp, 2006, p. 540; see also Lyster, 1998), thus precluding clear comparisons between recasts and negotiation. Further questioning the feasibility of distinguishing recasts from negotiation for meaning, specifically in classroom settings, Ellis (2001) called the distinction "an uncomfortable one, because it necessitates interpreting the intention of the teacher" (p. 24).

The comparison of recasts, explicit correction, and prompts is not only feasible but also has the potential to contribute theoretically to the study of SLA, because of the different types of evidence they afford and the different types of processing they activate. As Philp, Mackey, and

Oliver (2008) noted, "interaction research has continued to develop from roots found in first language acquisition research on child-directed speech" (p. 4). Researchers drawing on first language (L1) acquisition studies such as those by Farrar (1990, 1992) advocate the use of recasts, arguing that they (a) serve as both positive and negative evidence and (b) maintain a primary focus on meaning while enabling learners to notice errors in their interlanguage production (Doughty, 2001; Long 1996, 2007). However, as Ellis and Sheen (2006) suggested, recasts can serve as both positive and negative evidence if learners perceive the teacher's corrective intention but only as positive evidence if learners are not consciously aware of their illocutionary force (see also Leeman, 2003). In classroom settings in which there is already a primary focus on meaning (e.g., content-based instruction), it is difficult for learners to know whether teachers are responding to the veracity of their utterance or its form (Lyster, 1998, 2007).

Other CF researchers hypothesize that L2 learning resembles skill acquisition, which entails a gradual transition, through practice in meaningful contexts, from effortful to automatic use of rules, and that L2 instruction should be designed to help L2 learners develop automaticity in target language use (see DeKeyser, 2003, 2007). These researchers advocate CF such as prompts, which provide learners with instances of negative evidence combined with cues leading to the retrieval of alternative forms, thus allowing for opportunities to practice emergent target forms in contexts of interaction (e.g., Lyster, 2007; Ranta & Lyster, 2007). Long (2007) questioned the psycholinguistic rationale for prompting and claimed that "acquisition of new knowledge is the major goal, not 'automatizing' the retrieval of existing knowledge" (p. 102), but, as Lyster pointed out, "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" (p. 119).

Age Factors

Feedback effectiveness as a function of learners' age has, surprisingly, been given scarce research attention despite the well-known impact of age on L2 development. Notable exceptions are studies by Oliver (2000) and Mackey and Oliver (2002). In her classroom, observational study (10 adult and 10 child English-as-a-second-language [ESL] classrooms), Oliver discovered that teachers responded differently to adult and child learners' nontargetlike production in the course of oral interaction, noting that "teachers have greater expectations for adult learners, encouraging greater risk taking from their learners" (p. 138), whereas with

children, teachers "hold greater control of the interactions, reducing the opportunity for nontargetlike utterances by the younger learners" (p. 138). Oliver concluded that CF might be a good fit for child learners because it "is provided at a time in their cognitive and psycholinguistic development when it is most conducive for their learning" (p. 143). Mackey and Oliver followed up with a pretest and posttest experiment in which 22 child ESL learners participated in three 30-min information-gap tasks in teacher-student dyads. With results showing that child learners benefited from interactional CF and that the effects were more immediate than those observed with adult ESL learners (cf. Mackey & Philp, 1998), Mackey and Oliver concluded that children's sensitivity to recasts seems to be due to the fact that implicit feedback such as recasts is functionally similar to L1 feedback given by caregivers.

THE PRESENT STUDY

Inspired by this review of relevant CF studies, we conducted a quantitative research synthesis to summarize and interpret the findings obtained in classroom quasi-experimental studies of CF. Our intention is to achieve a better understanding of CF effectiveness, with a focus on its pedagogical capability, and to reveal relevant patterns across studies that will generate worthwhile research questions for future CF studies designed to investigate the complex of variables that affect CF effectiveness in L2 classrooms. Unlike a narrative review of CF effectiveness, the meta-analysis approach adopted here to synthesize previous studies entails a statistical procedure that permits meaningful comparisons of their multifaceted results.

The usefulness of meta-analysis as a trustworthy tool for research synthesis has been widely recognized in the area of instructed SLA studies since Norris and Ortega's (2000) seminal study (see also Norris & Ortega, 2006). A meta-analysis involves "gathering and weighting available evidence offered by results from all primary studies addressing a common research problem" (Norris & Ortega, 2000, p. 423) by calculating Cohen's *d* index, which is designed to measure an effect size between two contrastive groups of means (i.e., pretest and posttest for time and experimental vs. control groups for group). As Cohen (1988) explained, "the larger this value, the greater the extent to which the phenomenon under study is manifested" (p. 10).

The current meta-analysis adopts inclusion and exclusion criteria that differ from those adopted by Russell and Spada (2006) and Mackey and Goo (2007) to answer not only questions of theoretical relevance but also questions of classroom applicability. In other words, rather than taking into account all oral CF studies conducted in a variety of settings, only quasi-experimental studies conducted in L2 classrooms

were selected. Due to these focused selection criteria, only 15 studies are included in the present meta-analysis (N = 827). Therefore, results will need to be interpreted with caution even if they may be considered sufficiently reliable to have implications for future CF studies and for practitioners in L2 classrooms, given the 95% confidence intervals calculated for the purpose of robust statistical analyses. Our analyses take into account a variety of independent and dependent variables conducted in primary research (i.e., CF types, immediate vs. delayed posttest results, and types of outcome measurements) and compare the relative effectiveness of CF in L2 classrooms with contextual factors (i.e., second language [SL] vs. foreign language [FL] settings), length of treatment, and learners' age (i.e., child vs. young adult vs. adult learners). The research questions are formulated as follows:

1. How effective is CF on target language development in L2 classrooms?
2. If effective, to what extent does CF effectiveness vary according to the following variables?
   a) Types of CF
   b) Types and timing of outcome measures
   c) Instructional setting (SL vs. FL classroom)
   d) Length of treatment (brief vs. short-to-medium vs. long)
   e) Learners' age

Considering that the selection and characterization of the reviewed studies need to be clarified for the purpose of a robust meta-analysis, the procedures of inclusion versus exclusion will now be outlined.

METHOD

Data Selection

The current study focuses exclusively on published work without including so-called fugitive literature (e.g., unpublished doctoral dissertations and conference presentations). The fact that studies accepted for publication tend to have statistically significant results gives rise to the so-called file-drawer problem whereby studies without statistical significance remain unpublished (see Lipsey & Wilson, 2001, for details). However, following Norris and Ortega (2000), we did not attempt to retrieve any so-called fugitive literature because our purpose is to investigate the current state of knowledge based on accessible published literature. Ultimately, this approach allows us to achieve "an accurate synthesis of exactly those findings from those studies that are published and reported, and that therefore in many ways define this research domain" (Norris & Ortega, p. 432).

We searched electronic databases such as the Education Resources Information Center (ERIC) and the Linguistic and Language Behavior Abstracts (LLBA) to identify studies of oral CF published since 1980 in major academic journals (e.g., *Applied Linguistics*, *The Canadian Modern Language Review*, *Language Awareness*, *Computer Assisted Language Learning*, *Language Learning*, *The Modern Language Journal*, *Studies in Second Language Acquisition*, *System*, *TESOL Quarterly*), including this special issue of *SSLA* devoted to CF. The following key words, among others, were used: *feedback, CF, explicit feedback, implicit feedback, recasts, prompts, explicit correction, negotiation, L2 learning,* and *pre/post design*. We also referred to relevant review chapters to check potential sources (e.g., Doughty & Long, 2003; Doughty & Williams, 1998; Ellis, 1999, 2003; Long, 2007; Lyster, 2007; Mackey, 2007). After excluding observational studies that did not experimentally measure learners' interlanguage development (e.g., Ellis, Basturkmen, & Loewen, 2001; Lyster & Ranta, 1997; Oliver, 2000) as well as experimental and quasi-experimental studies that adopted a pretest and posttest design but did not report necessary information for calculating Cohen's *d* index (e.g., Doughty & Varela, 1998; Havranek & Cesnik, 2001; Loewen, 2005; Loewen & Philp, 2006), 34 CF studies qualified for a meta-analysis (see Table 1; note that two of the laboratory-based studies are computer-assisted language learning studies in which the feedback appeared in written format on the computer screen yet was provided during computer-mediated communication designed to simulate face-to-face communication).[2]

We then excluded all laboratory-based studies (e.g., Carroll & Swain, 1993; Leeman, 2003; Lyster & Izquierdo, 2009; Mackey & Philp, 1998; McDonough, 2005, 2007) and selected only quasi-experimental classroom-based studies. Although it would have been possible to conduct a meta-analysis of 34 CF studies, we decided against conflating research settings. As Norris and Ortega (2006) claimed, "the problem is to define what makes studies 'similar enough' to be included in a review, or 'different enough' to be excluded, the so-called *apples and oranges* problem" (p. 16). With respect to CF studies, one of the controversial problems that influence the construct validity of primary research is whether studies were conducted in classroom or laboratory settings. Thus, interpreting data extracted from both classroom- and laboratory-based studies together does not correspond to the main purpose of this study, which is to examine CF effectiveness in L2 classrooms. It should be noted, however, that the instructional treatments employed in the 15 studies differ in substantial ways; Table 2 provides an overview of the 15 studies with information concerning number of participants, their age and L1 background, linguistic targets, CF types, treatment length, and types of outcome measures.

Table 1. Thirty-four classroom- and laboratory-based CF studies qualified for a meta-analysis

Classroom-based studies (N = 15)	Laboratory-based studies (N = 19)
Ammar and Spada (2006)	Carroll and Swain (1993)
DeKeyser (1993)	Carroll, Swain, and Roberge (1992)
Ellis (2007)	Han (2002)
Ellis, Loewen, and Erlam (2006)	Inagaki and Long (1999)
Ellis, Rosszell, and Takashima (1994)	Ishida (2004)
Herron (1991)	Iwashita (2003)
Herron and Tomasello (1988)	Kim and Mathes (2001)
Loewen and Nabei (2007)	Leeman (2003)
Lyster (2004)	Loewen and Erlam (2006)
Mackey (2006)	Long, Inagaki, and Ortega (1998)
Muranoi (2000)	Lyster and Izquierdo (2009)
Sheen (2007)	Mackey and Oliver (2002)
Takashima and Ellis (1999)	Mackey and Philp (1998)
Tomasello and Herron (1989)	McDonough (2005)
Yang and Lyster (this issue)	McDonough (2007)
	McDonough and Mackey (2006)
	O'Relly, Flaitz, and Kromrey (2001)
	Sagarra (2007)
	Sauro (2009)

Data Computation

To calculate Cohen's *d* index, three statistical elements are required: (a) group means, (b) standard deviations, and (c) sample sizes. The equation in (1), adapted from Cohen (1988), was used, with the pooled between-groups standard deviation.[3]

$$d = \frac{M_1 - M_2}{\sigma_{\text{pooled}}}$$
$$\sigma_{\text{pooled}} = \frac{(n_1 - 1)\sigma_1 + (n_2 - 1)\sigma_2}{(n_1 - 1) + (n_2 - 1)} \quad (1)$$

where M is the mean and σ is the standard deviation.

When the published study provided only t or F values, the formula provided in (2) was used to obtain d values (Lipsey & Wilson, 2001; see, e.g., Herron, 1991; Takashima & Ellis, 1999).

$$d = \frac{t\sqrt{(n^1 + n^2)}}{\sqrt{(n^1 + n^2)}}, \quad d = \frac{\sqrt{F}\sqrt{(n^1 + n^2)}}{\sqrt{(n^1 + n^2)}}. \quad (2)$$

When the study reported proportions of participants who showed improvements, Cohen's *d* was calculated based on arcsine transformation (Lipsey & Wilson; see, e.g., Mackey, 2006): Cohen's $d = \text{arcsine}_{\text{treatment}} - \text{arcsine}_{\text{control}}$.

The current study calculated 95% confidence intervals to analyze whether two groups of effect sizes are significantly different (e.g., immediate vs. delayed posttests, recasts vs. prompts). Cohen's *d* can be identified by calculating two contrastive means of dependent variables. Because Norris and Ortega (2006) recommended analyzing *d* values separately for between- and within-group comparisons in instructed SLA studies, to examine the data in a thorough manner, *d* values were calculated separately for the mean difference between experimental and control groups at the time of posttests (between group; $n = 15$) as well as between pretests and posttests (within group; $n = 10$). Unlike previous meta-analysis studies (i.e., Mackey & Goo, 2007; Norris & Ortega, 2000), all 15 CF studies in the current meta-analysis included control groups; therefore, no baseline groups (i.e., using least attention-focused groups when there is no control group) for between-group differences had to be chosen, which enabled an independent analysis of between- and within-group contrasts. Although some scholars suggest that one study should contribute to only one effect size (Lipsey & Wilson, 2001), we decided to implement inclusive adoption of multiple independent and dependent variables (e.g., treatment groups, CF types, timing of posttests, and outcome measures) to conduct a fine-grained analysis of all developmental changes among students in any given study, especially in light of the complexities of L2 classrooms.[4]

Coding

This section explains how the independent variables were operationalized, including four instructional factors (i.e., types of CF, instructional setting, treatment length, and participants' age) and two methodological factors (i.e., types and timing of outcome measurements). CF intensiveness (i.e., whether the study was designed to direct CF intensively at a specific linguistic target) was not treated as an independent variable because of the number of targets that the primary studies focused on varied, and it was difficult to define intensiveness versus extensiveness categorically in this meta-analysis.[5]

Types of CF. Following Lyster and Ranta's (1997) original taxonomy, CF types were classified as (a) recasts, (b) explicit correction, and (c) prompts (what Lyster and Ranta called *negotiation of form*: i.e., clarification requests, repetition of error, elicitation, metalinguistic clues). Seven studies were selected for recasts (i.e., Ammar & Spada, 2006;

Table 2. Overview of the 15 selected studies

Studies	*N*	Age	L1 background	Target forms (and target language)	CF types	Length	Test types[a]
Second Language settings							
DeKeyser (1993)	35	17	Dutch	Morphosyntax (French)	Explicit correction	Long	FR, CR
Lyster (2004)	179	10.5	English	Grammatical gender (French)	Recasts versus prompts	Long	FR, CR, SR
Ammar and Spada (2006)	64	11.5	French	Possessive determiners (English)	Recasts versus prompts	Long	FR, MJ
Mackey (2006)	28	24.2	Various	Question forms, plurals, past tense (English)	Recasts and prompts	Short-to-medium	FR
Ellis et al. (2006)[b]	32	25	East Asian (77%)	Regular past tense (English)	Recasts versus prompts	Short-to-medium	CR, MJ
Ellis (2007)[b]	32	25	East Asian (77%)	Regular past tense and comparative *–er* (English)	Recasts versus prompts	Short-to-medium	CR, MJ
Sheen (2007)	80	24–27[c]	Various	Definite and indefinite articles (English)	Recasts versus explicit correction	Short-to-medium	CR, MJ

Foreign Language settings							
Herron and Tomasello (1988)	32	19.5	English	Transfer errors (French)	Explicit correction	Brief	CR
Tomasello and Herron (1989)	32	19	English	Overgeneralization errors (French)	Explicit correction	Brief	CR
Herron (1991)	25	19.5	Japanese	Overgeneralization errors (French)	Explicit correction	Brief	CR
Ellis et al. (1994)	61	19	Japanese	Dative alternation (English)	Explicit correction	Brief	MJ
Takashima and Ellis (1999)	61	19	Japanese	Regular and irregular past tense (English)	Prompts	Short-to-medium	FR
Muranoi (2000)	91	18	Japanese	Indefinite articles (English)	Recasts and prompts	Short-to-medium	FR, CR, MJ
Loewen and Nabei (2007)	35	19	Japanese	Question forms (English)	Recasts versus prompts	Brief	MJ
Yang and Lyster (this issue)	72	20	Chinese	Regular and irregular past tense (English)	Recasts versus prompts	Short-to-medium	CR

[a]The abbreviation FR corresponds to free constructed-response measures, CR corresponds to constrained constructed-response measures, SR corresponds to selected-response measures, and MJ corresponds to metalinguistic judgments.
[b]These are the same participants.
[c]Because Sheen (2007) did not report the average mean of learners' age, we contacted her and recalculated their age.

Ellis, 2007; Ellis, Loewen, & Erlam, 2006; Loewen & Nabei, 2007; Lyster, 2004; Sheen, 2007; Yang & Lyster, this issue), seven studies for prompts (i.e., Ammar & Spada; Ellis; Ellis et al.; Loewen & Nabei; Lyster; Takashima & Ellis, 1999; Yang & Lyster), and six for explicit correction (i.e., DeKeyser, 1993; Ellis, Rosszell, & Takashima, 1994; Herron, 1991; Herron & Tomasello, 1988; Sheen; Tomasello & Herron, 1989). Although Muranoi (2000) labeled two of his treatment groups as recasts, both groups were excluded from our analysis of CF types because they received not only recasts but also requests for repetition (i.e., clarification requests, thus a combination of recasts and prompts). Mackey's (2006) study was also excluded from the comparisons of CF types because the feedback treatment in her study conflated both recasts and negotiation (which included clarification requests that prompted participants to self-repair). The studies by Ellis and Ellis et al. were framed as comparisons of explicit and implicit feedback but operationalized explicit CF as metalinguistic information in the form of a prompt (e.g., *You need past tense*) rather than as explicit correction (e.g., "*No, not goed—went*," Ellis et al., p. 341).

Referring specifically to these studies, Figure 1 attempts to group CF types along a continuum that ranges from implicit to explicit and according to the dichotomous distinction between reformulations and prompts. Although we place recasts toward the implicit end of the continuum, we do so in relation to explicit correction for the purpose of distinguishing these two types of reformulation while acknowledging that recasts themselves range from more implicit to more explicit (Loewen & Philp, 2006; Sheen, 2006). To classify prompts as implicit or

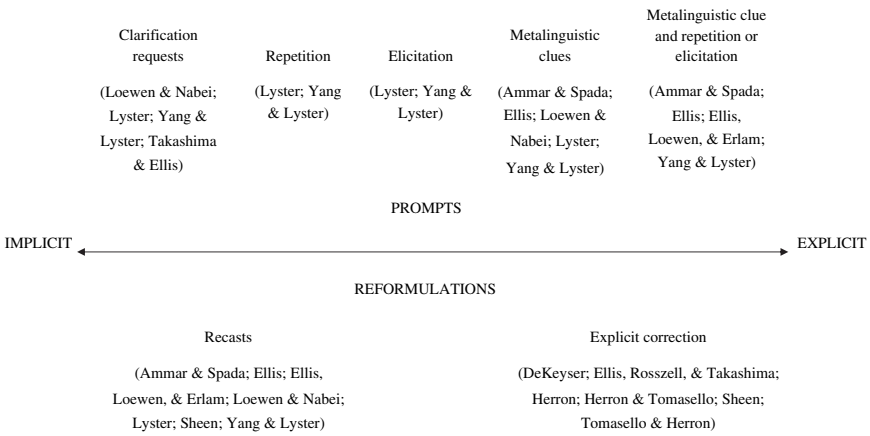

Figure 1. Types of CF.

explicit, we followed suggestions by Ellis (2006), Loewen and Nabei (2007), and Lyster (2002) that clarification requests and repetition are more implicit than elicitation and metalinguistic clues. The classification of prompts at discrete points along this continuum is rather crude and open to further refinement, but we reiterate that the distinguishing feature of prompts in this classification pertains less to their varying degrees of implicitness or explicitness and more to their common trait of withholding correct reformulations.

Whereas recasts and prompts in the studies included in this meta-analysis were generally delivered in the context of communicatively driven teacher-student interaction, explicit correction was operationalized with slightly less consistency and across a broader spectrum of instructional activities. In Sheen's (2007) study, explicit correction, provided during narrative-retelling tasks, included the correct form along with metalinguistic explanation—for instance, *The fox. You should use the definite article* ***the*** *because you've already mentioned* ***fox***. The explicit correction treatment in DeKeyser's (1993) study lasted an entire school year and was thus arguably less likely to maintain the consistent use of only one CF type, as illustrated by the only example provided (*How many times?*) that serves more as a prompt. Yet the teacher in the treatment class had been asked "to correct mistakes as frequently and explicitly as possible" (p. 505) and her use of CF was observed to be "usually quite conspicuous" (p. 506). In contrast, the comparison teacher was asked to provide no feedback, yet occasionally provided "inconspicuous" corrections "without explicitly saying 'no', 'wrong', or anything of the kind, and without making the students self-correct" (p. 506). Explicit correction in the studies by Ellis et al. (1994), Herron (1991), Herron and Tomasello (1988), and Tomasello and Herron (1989) was not delivered during communicative tasks per se but rather through so-called garden-path techniques that induced learners to make errors in their use of target forms. In response to the errors, the teacher wrote the incorrect form on the chalkboard, crossed it out, wrote the correct form, and then said it aloud before providing a brief explanation.

Durability. The durability of the effects of form-focused instruction and interaction has been difficult to ascertain in previous SLA research. On the one hand, Norris and Ortega (2000) showed that the effectiveness of L2 instruction appeared to decrease gradually between immediate and delayed posttests, but they included various types of instruction (e.g., focus-on-form, focus-on-forms, and focus-on-meaning instruction). On the other hand, in their meta-analysis of the effectiveness of L2 interaction only, Mackey and Goo (2007) observed that "the efficacy of interaction on learner performance in both delayed tests is significantly greater than in the immediate post-tests" (p. 425). In the

present study, effect sizes were calculated separately for immediate and delayed-posttest sessions, with posttests conducted within 1 week of an instructional treatment considered as immediate, and those after 2 weeks and up to 6 weeks considered delayed.

Types of Outcome Measures. Because different types of dependent-variable measures assess learners' interlanguage development from different perspectives, their effects are expected to vary in size. Following Norris and Ortega's (2000) classification, CF effects were examined in relation to four different types of outcome measure:[6]

1. *Free constructed-response measures* require learners to produce the target language freely without many constraints.
2. *Constrained constructed-response measures* require learners to complete tasks in which the use of the target features was necessary.
3. *Selected-response measures* require learners to select the correct answer among several alternatives.
4. *Metalinguistic judgments* require learners to judge the grammaticality of target structures.

Two previous meta-analysis studies provided conflicting results about the influence of outcome measures on L2 performance. Whereas Norris and Ortega (2000) found that effect sizes associated with metalinguistic judgments and free constructed-response measures were substantially lower than those with selected-response measures and constrained constructed-response measures, Mackey and Goo (2007) noted a linear relationship, on immediate posttests in particular: Closed-ended production was better than open-ended production, which was better than prompted response (i.e., a combination of metalinguistic judgments and selected-response measures).

Instructional Setting. Mackey and Goo's (2007) meta-analysis of interaction studies revealed differences between SL and FL settings, with significantly larger effects in FL settings (d = 0.88) than in SL settings (d = 0.64), but no difference at delayed posttests. In the present meta-analysis, the influence of FL versus SL settings on CF effectiveness is also compared, but rather than estimating the extent to which learners could access the target language outside of the classroom, we adopted Stern's (1983) definition to distinguish these two settings: "A 'second language' usually has official status or a recognized function within a country which a foreign language has not" (p. 16). Following this definition, seven studies were conducted in SL settings and eight in FL settings, as shown in Table 2.

Length of Treatment. Norris and Ortega's (2000) meta-analysis took into account the length of instructional treatment as an independent variable by analyzing the effects of brief treatments (less than 1 hr),

short treatments (1–2 hr), medium treatments (3–6 hr), and long treatments (more than 7 hr). They found that less instruction could actually be more effective. For our analysis, we considered five studies as brief treatments (less than 1 hr), seven studies as short-to-medium treatments (1–3 hr), and three studies as long treatments (more than 7 hr).

Learners' Age. To investigate the age-related effects of CF, all *d* values (based on CF types and timing of posttests) were compared with participants' age. Most studies reported the mean age of participants, but others reported only their current educational level, which enabled us to predict participants' age (e.g., sixth-grade students are 11.5 years old, first-year university students are 18 years old). The participants in the studies included in the meta-analysis fall into three age categories: (a) *Child learners* are students in elementary school with a mean age of 10–12 years (n = 243); (b) *young-adult learners* are students either in the last year of high school or a university undergraduate program and have a mean age of 17–20 years (n = 444); (c) *adult learners* are ESL students who attend a language school or community college and have a mean age above 23 years (n = 140). However, this study does not compare these as three discrete groups but instead considers age as a continuous variable in a simple regression analysis designed to examine negative linear relationships between the effects of CF and learners' age.

RESULTS

The results that concern the overall effectiveness of CF in L2 classrooms are presented first, followed by results that assess how its effectiveness varies according to CF types, timing and types of outcome measures, instructional setting, treatment length, and age. The 15 CF studies created 43 *d* values for between-group contrasts, and 10 studies yielded 33 *d* values for within-group contrasts. For the between-group contrasts, each study contributes an average of 2.86 *d* values (*SD* = 1.24): k (i.e., the number of instructional treatments that contribute effect sizes) = 13 for adult learners, k = 22 for young-adult learners, and k = 8 for child learners. For within-group differences, each study contributes an average of 3.3 *d* values (*SD* = 1.25): k = 12 for adult learners, k = 13 for young-adult learners, and k = 8 for child learners (see the Appendix for a summary of all results).

Effects of CF

The between-group contrasts for CF effects created 43 *d* values, with a mean effect size of 0.74 (*SD* = 0.47) and confidence intervals (CIs) above zero (0.58–0.90), which demonstrates that the impact of CF is significant

in L2 classrooms with medium-to-large effects. Within-group contrasts created 33 *d* values, with a mean effect size of 0.91 (CI = ± 0.15, *SD* = 0.46). Even though the control groups demonstrated a small effect size of 0.39 (CI = ± 0.09, *SD* = 0.20), CF effectiveness proved significant in the context of within-group contrasts because there is no overlap of confidence intervals between experimental groups that receive CF (0.76–1.06) and the control groups that did not (0.30–0.48). The effects of CF could be considered medium (*d* = 0.52) if the effect found for the control groups (*d* = 0.39) is subtracted from the overall large effect (*d* = 0.91). These results are presented in Figure 2.

Types of CF

With respect to the between-group contrasts, the results support medium effects of 0.53 for recasts (*n* = 7, *k* = 13, CI = ± 0.21, *SD* = 0.39), large effects of 0.83 for prompts (*n* = 7, *k* = 15, CI = ± 0.27, *SD* = 0.53), and large effects of 0.84 for explicit correction (*n* = 6, *k* = 10, CI = ± 0.27, *SD* = 0.27). In the within-group contrasts, we found medium effect sizes of 0.70 for recasts (*n* = 7, *k* = 13, CI = ± 0.16, *SD* = 0.31), large effect sizes of 1.14 for prompts (*n* = 6, *k* = 12, CI = ± 0.25, *SD* = 0.44), and medium effect sizes of 0.60 for explicit correction (*n* = 3, *k* = 4, CI = ± 0.37, *SD* = 0.37). Taken together, with confidence intervals above zero, recasts, prompts, and explicit correction are all significantly effective in L2 classrooms both in the between- and within-group contrasts. Whereas explicit correction

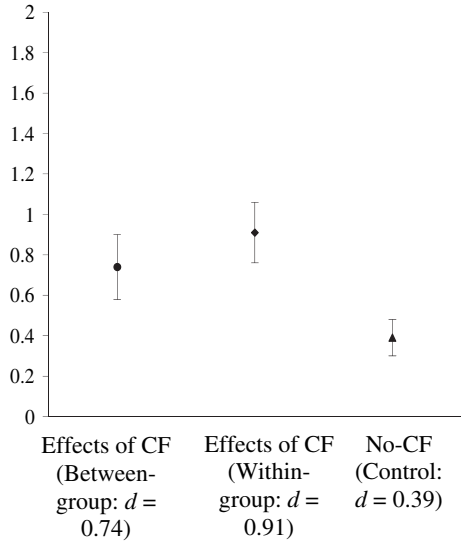

Figure 2. Effects of CF in L2 classrooms.

did not prove significantly different from either recasts or prompts—neither in between- nor within-group contrasts—the analysis of confidence intervals found the differential effectiveness of prompts to be significantly larger than that of recasts in the within-group contrasts (see Figure 3). Considered reliable, therefore, the analysis yields the following three patterns and relative weights associated with different types of CF in L2 classrooms:

1. Recasts, prompts, and explicit correction all yielded significant effects.
2. Prompts yielded large effect sizes and proved significantly more effective in the within-group contrasts than recasts.
3. Effects of explicit correction cannot be distinguished from those of recasts and prompts.

We acknowledge, however, that standard deviations and confidence intervals of all three types of CF (i.e., prompts, recasts, and explicit correction) are widely ranged. One possible reason for such variability could be that CF treatments are difficult to implement in classroom settings with the same kind of consistency found in more controlled situations such as laboratory settings.

Immediate and Delayed Effects of CF

Between-group contrasts generated 25 *d* values at immediate posttests with a mean effect size of 0.63 (CI = ± 0.18, *SD* = 0.47). Ten CF studies

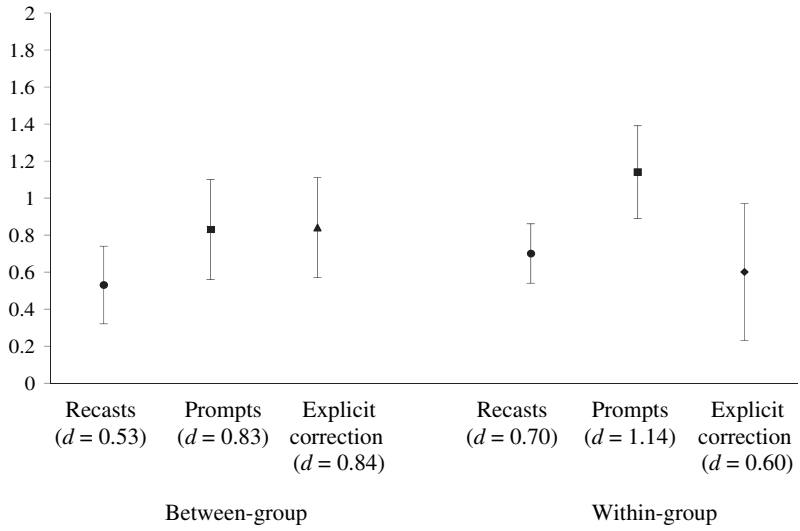

Figure 3. Types of CF in L2 classrooms.

generated 16 *d* values at delayed posttests, with a mean effect size of 0.84 (CI = ± 0.21, *SD* = 0.45). As for within-group contrasts, 10 studies generated 19 *d* values at immediate posttests with a mean effect size of 0.84 (CI = ± 0.21, *SD* = 0.47), whereas 8 studies generated 14 *d* values at delayed posttests, with a mean effect size of 1.01 (CI = ± 0.23, *SD* = 0.44). Both immediate and delayed effects of CF thus exhibit medium-to-large effect sizes, with overlapping confidence intervals as illustrated in Figure 4, which indicates that there is no significant decline in gains between immediate and delayed posttests.

Types of Outcome Measures

Three types of outcome measures were considered for this analysis (i.e., free constructed-response measures, constrained constructed-response measures, and metalinguistic judgments), which yielded 78 *d* values for between-group contrasts (*n* = 15) and 69 *d* values for within-group contrasts (*n* = 10).[7] On average, one study uses 1.93 outcome measures (*SD* = 0.96) to examine the effectiveness of an instructional treatment (i.e., the number of tests used in the primary studies roughly ranges between 0.96 and 2.89). With respect to the between-group contrasts, the mean effect size of free constructed-response measures is 0.97 (*n* = 5, *k* = 13, CI = ± 0.29, *SD* = 0.51), the mean effect size of

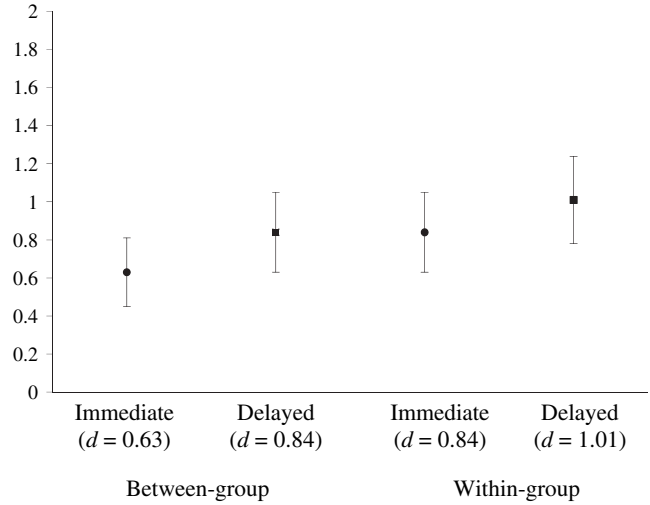

Figure 4. Immediate and delayed effects of CF.

constrained constructed-response measures is 0.70 (n = 9, k = 38, CI = ± 0.15, SD = 0.48), and the mean effect size of metalinguistic judgments is 0.45 (n = 5, k = 27, CI = ± 0.19, SD = 0.52). With respect to within-group contrasts, we identified the mean effect size of free constructed-response measures as 1.25 (n = 3, k = 9, CI = ± 0.38, SD = 0.58), the mean effect size of constrained constructed-response measures as 0.86 (n = 6, k = 29, CI = ± 0.14, SD = 0.39), and the mean effect size of metalinguistic judgments as 0.70 (n = 6, k = 31, CI = ± 0.21, SD = 0.60).

The means indicate large effect sizes for free constructed-response measures and medium-to-large effect sizes for constrained constructed-response measures and metalinguistic judgments. The analysis of confidence intervals shows that the effects of free constructed-response measures are significantly larger than those of metalinguistic judgments in the between-group contrast, and the difference between these two measures almost reached significance, with slight overlapping confidence intervals in the within-group contrasts. Therefore, students in L2 classrooms appear to exhibit improvement more clearly in free constructed-response measures than in metalinguistic tasks. Due to overlapping confidence intervals, however, the differential effects of constrained constructed-response measures in comparison with free constructed-response measures and metalinguistic judgments remain unclear. The results are graphically presented in Figure 5, and d values are summarized in Table 3.

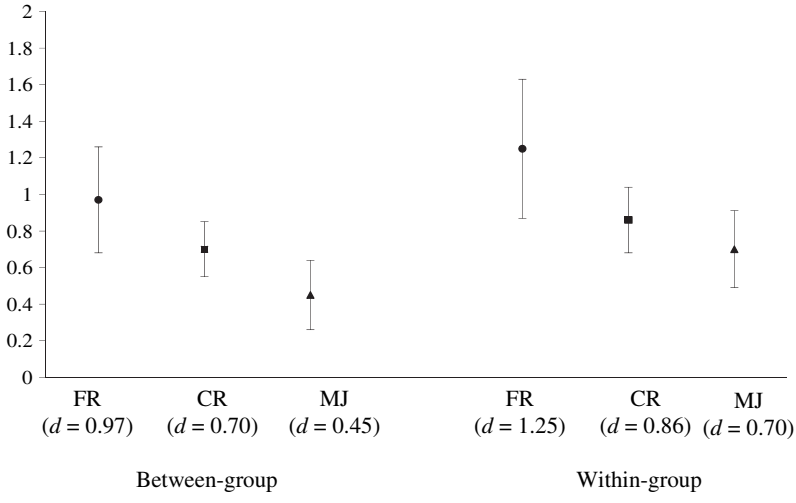

Figure 5. Types of outcome measures.

Table 3. Summary of *d* values in 15 selected studies

Variables	*n*	*k*	*M*	95% CI Lower	95% CI Upper	*SD*
A. *Between-Group*						
General effects of CF	15	43	0.74	0.58	0.86	0.47
CF types						
Recasts	7	13	0.53	0.32	0.74	0.39
Prompts	7	15	0.83	0.56	1.10	0.53
Explicit correction	6	10	0.84	0.57	1.11	0.27
Timing of posttests						
Immediate	15	25	0.63	0.45	0.81	0.47
Delayed	10	18	0.84	0.63	1.05	0.45
Settings						
SL	7	22	0.74	0.54	0.94	0.48
FL	8	21	0.70	0.51	0.89	0.45
Length						
Brief (<1 hr)	5	10	0.72	0.36	1.08	0.58
Short-medium (1–3 hr)	7	24	0.57	0.44	0.70	0.33
Long (>7 hr)	3	9	1.13	0.82	1.44	0.47
Types of outcome measures						
FR	5	13	0.97	0.68	1.33	0.51
CR	9	38	0.70	0.55	0.85	0.48
MJ	5	27	0.45	0.26	0.64	0.52
B. *Within-Group*						
General effects of CF	10	33	0.91	0.76	1.06	0.46
Control (no CF)	10	17	0.39	0.30	0.48	0.20
CF types						
Recasts	7	13	0.70	0.54	0.86	0.31
Prompts	6	12	1.14	0.89	1.39	0.44
Explicit correction	3	4	0.60	0.23	0.97	0.37
Timing of posttests						
Immediate	10	19	0.84	0.63	1.05	0.47
Delayed	8	14	1.01	0.78	1.24	0.44
Settings						
SL	6	21	0.89	0.68	1.10	0.49
FL	4	12	0.96	0.73	1.19	0.41
Length						
Short-medium (1–3 hr)	5	20	0.87	0.68	1.06	0.44
Long (>7 hr)	3	9	1.11	0.75	1.44	0.55
Types of outcome measures						
FR	3	9	1.25	0.85	1.63	0.58
CR	6	29	0.86	0.72	1.00	0.39
MJ	6	21	0.70	0.49	0.91	0.60

Instructional Setting

To investigate how the effects of CF are related to instructional setting, effect sizes were calculated separately for studies conducted in SL and FL settings. As for the between-group contrasts, the mean effect size for SL studies is 0.74 (n = 7, k = 22, CI = ± 0.20, SD = 0.48) and 0.70 (n = 8, k = 21, CI = ± 0.19, SD = 0.45) for FL studies. Within-group contrasts yielded mean effect sizes of 0.89 (n = 6, k = 21, CI = ± 0.21, SD = 0.49) for SL studies and of 0.96 (n = 4, k = 12, CI = ± 0.23, SD = 0.41) for FL studies. This analysis does not indicate different CF effects in SL versus FL settings because of the overlaps both in the between- and the within-group contrasts, as illustrated in Figure 6.

Length of Treatment

With respect to between-group contrasts, the mean effect size of brief treatments is 0.72 (n = 5, k = 10, CI = ± 0.36, SD = 0.58), the mean effect size of short-to-medium treatments is 0.57 (n = 7, k = 24, CI = ± 0.13 SD = 0.33), and the mean effect size of long treatments is 1.13 (n = 3, k = 9, CI = ± 0.31, SD = 0.47). With respect to within-group contrasts, we identified the mean effect size of short-to-medium treatments as 0.87 (n = 5, k = 20, CI = ± 0.19, SD = 0.44) and of long treatments as 1.11 (n = 3, k = 9, CI = ± 0.36, SD = 0.55).[8] Taken together, these results reveal the following

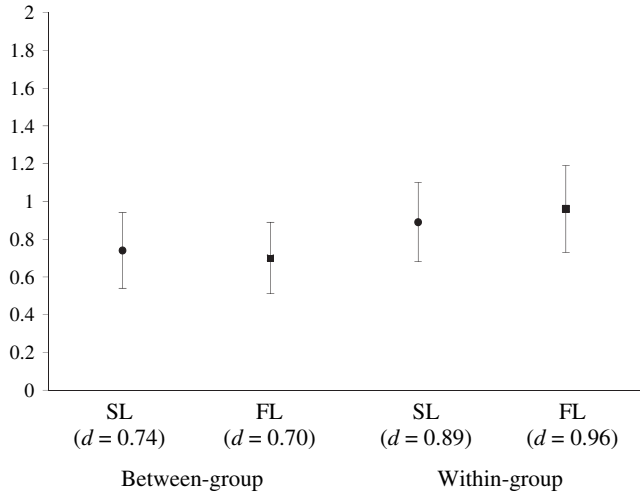

Figure 6. Effects of CF in SL and FL settings.

pattern: Whereas the value of brief treatments remains unclear, the analysis of confidence intervals shows that long treatments (0.82–1.44) are significantly more effective than short-to-medium treatments (0.44–0.70) in the between-group contrasts (see Figure 7).

Age

To investigate how the effects of CF are linearly related to learners' age, instead of calculating *d* values for each group, we chose to perform a simple regression analysis for all *d* values in the between-group contrasts (n = 15, k = 43) and the within-group contrasts (n = 10, k = 33), by comparing the average mean of all *d*-values as a dependent variable (an outcome) and the participants' age as an independent variable (a predictor factor). Although the data need to be interpreted with caution due to the small number of child and adult CF studies, our purpose here is to identify possible patterns of age-related effects on CF effectiveness worthy of further investigation.

The between-group contrasts yielded a significant linear relationship between effects of CF and age, $F(1, 41) = 14.573$, $p < .0001$. The model used here shows that the age factor explains 26.2% of the variance of the effectiveness of CF in L2 classrooms (r = –.51). With respect to within-group contrasts, a significant linear relationship was identified

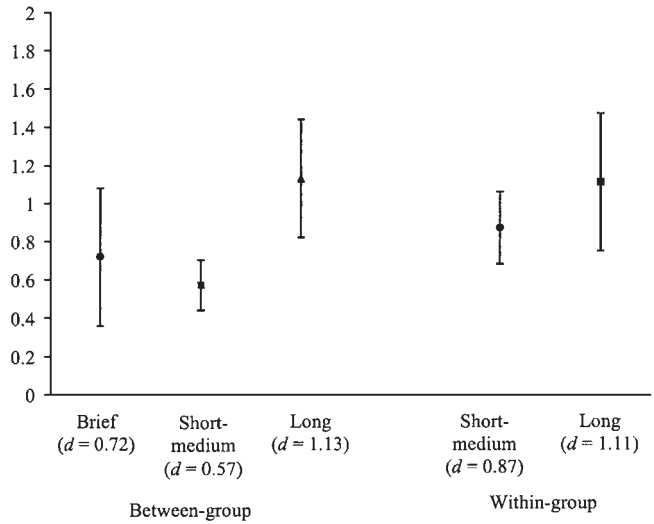

Figure 7. Effects of CF and length of instruction.

between effects of CF and the age factor, $F(1, 31) = 6.075$, $p < .02$. The model used here shows that the age factor explains 16.4% of the variance of the effectiveness of CF in L2 classrooms ($r = -.40$). Thus, general effects of CF and learners' age are negatively correlated, and the possible pattern for effects of CF is such that younger learners are more affected than older learners.[9]

DISCUSSION

This discussion addresses our research questions, which asked whether CF is effective in L2 classrooms and, if so, whether its effectiveness varies according to (a) types of CF, (b) types and timing of outcome measures, (c) instructional settings, (d) treatment length, and (e) learners' age. However, it is important to emphasize the need to interpret the results with caution and to consider them tentative, given the obvious limitations: the small *N*, the relatively broad range of confidence intervals and standard deviations in all contexts, and the inclusive contribution of several *k* from one sample study.

Overall Effects of CF

Based on the premise that results of laboratory-based studies cannot predict the pedagogical effectiveness of CF, the current meta-analysis included only classroom quasi-experimental CF studies to assess the pedagogical value of CF techniques in L2 classrooms. First, the current study shows that CF makes a significant impact on L2 learners' performance with a medium yet substantial effect ($d = 0.74$) for posttests in comparison with control groups (i.e., the between-group contrasts). This result concurs with those of previous meta-analysis studies (Mackey & Goo, 2007; Russell & Spada, 2006). Whereas students who received CF displayed large effect sizes ($d = 0.91$) in comparison with their pretest performance (i.e., the within-group contrasts), the present study also found that control groups exhibited medium effect sizes ($d = 0.39$). This result could be attributed to test-retest effects or to the fact that even participants in control groups, by virtue of being in classroom settings, also received instruction but without intentionally designed CF treatments. The results demonstrate that CF groups are significantly different from no-CF groups, with no overlap in confidence intervals. Although some scholars doubt the effectiveness of oral CF in L2 classrooms (e.g., Truscott, 1999), the current study, along with previous meta-analyses, illustrates its effectiveness.

CF Types

The analysis of differential effects of recasts, prompts, and explicit correction in L2 classrooms revealed three patterns: (a) Recasts, prompts, and explicit correction all yielded significant effects; (b) prompts yielded large effect sizes and proved significantly more effective in the within-group contrasts than recasts, which yielded medium effect sizes; and (c) the relative effects of explicit correction remained indistinguishable from those of recasts and prompts.

Therefore, whereas all three types of CF are able to positively affect L2 learners' interlanguage development, CF in classroom settings may be more effective when its delivery is more pedagogically oriented (i.e., prompts) than conversationally oriented (i.e., recasts). In theoretical terms, classroom learners appear to benefit from the positive evidence available in recasts as well as from the opportunities they provide to infer negative evidence, but these learners seem to benefit even more from the negative evidence available in prompts and from the greater demand they impose for producing modified output. That the effects of explicit correction cannot be distinguished from the effects of recasts and those of prompts might be attributable to the types of linguistic evidence available in explicit correction and the overlaps in this regard with recasts and prompts. In other words, explicit correction—similar to recasts—conveys positive evidence by providing the correct form; at the same time, explicit correction—similar to prompts—conveys negative evidence by indicating that the student's utterance was incorrect. Worthy of further pursuit in this regard would be research designed to tease apart the specific attributes of recasts, prompts, and explicit correction that contribute to their variable effectiveness in classroom settings.

Our results are different from those of Mackey and Goo (2007), who found much larger effects for recasts ($d = 0.96$), but we attribute this to differences in the inclusion criteria between the two studies: Our study included only classroom studies ($n = 15$) in the meta-analysis, whereas Mackey and Goo included CF studies conducted in both laboratory ($n = 15$) and classroom settings ($n = 5$). In short, the research setting factor considerably influences the relative effectiveness of different types of CF.

Durability

The overall durability of the effectiveness of CF was analyzed by contrasting effect sizes at immediate and delayed posttests. The

between- and within-group contrasts alike yielded medium-to-large effect sizes for both immediate and delayed posttests, with an overlap of confidence intervals, which shows that CF has an immediate impact on L2 performance and that the magnitude of the effectiveness does not decline significantly during the interval between the CF treatment and delayed posttests. We attribute these results to the strategic and intensive delivery of CF across the studies analyzed. In other words, in all but one study (i.e., DeKeyser, 1993), CF was provided during oral interaction that had been pedagogically designed to elicit target forms from students and to create strategic opportunities for teachers to provide CF. It may be the case that when CF is provided in this way, with relative intensity during form-focused activities designed to elicit target forms in the context of oral interaction, its effects are sufficiently robust not only to appear immediately after the instructional treatment but also to be sustained over time, which in the present analysis ranged from 2–6 weeks.

Outcome Measures

In their meta-analysis of L2 instructional interventions, Norris and Ortega (2000) noted that many studies employed outcome measures that favor the effectiveness of explicit treatments because these measures required "the application of explicit declarative knowledge under controlled conditions, without much requirement for fluent, spontaneous use of contextualized language" (p. 486), a research bias criticized by Doughty (2003). Mackey and Goo's (2007) meta-analysis of interaction studies as well as the present meta-analysis of CF studies (see also Spada & Tomita, in press) noted a tendency for more studies than those noted by Norris and Ortega to employ measures that require learners to produce target structures, with or without constraints, more than selective-response measures or grammatical judgment tasks.

This analysis of the relationship between CF effects and different outcome measures (i.e., free constructed-response measures, constructed-response measures, and metalinguistic judgments tasks) identified relatively large effect sizes for free constructed-response measures and medium-to-large effect sizes for constructed-response measures and metalinguistic judgments. Additionally, the analysis of the confidence intervals illustrated that free constructed-response measures generate substantially larger effects than metalinguistic judgments. The finding that the effects of oral CF are larger when students' improvement is tested by formats with fewer constraints, such as free constructed-response measures, is especially noteworthy. One possible interpretation of this finding is related to the principle of transfer-appropriate processing,

according to which the kind of cognitive processing that occurs during learning tasks should ideally resemble the kind of processing involved during communicative language use (e.g., Lightbown, 2008; Segalowitz, 1997, 2000). In other words, because CF is provided during actual target language use, its impact on learners' subsequent target language use in contexts "with relatively few constraints and with meaningful communication as the goal for L2 production" (Norris & Ortega, 2000, p. 440) is more transfer appropriate than the effects of decontextualized grammar lessons or communicative language teaching without CF. However, such explanations are speculative, due to the small *N* and relatively large confidence intervals and standard deviations. The tentative pattern suggested by the present analysis (i.e., larger effects for free constructed-response measures than for constructed-response measures and metalinguistic judgments) justifies further research into the variable effects of CF as measured by production tasks that are more or less controlled as well as by other tasks designed to assess both implicit and explicit knowledge of the target language.

SL Versus FL Settings

With respect to contextual influences on the effects of CF, no significant difference was found between studies in SL and in FL settings. This finding suggests either that distinctions between SL and FL settings are too fluid to yield distinguishable results (see Berns, 1990; Block, 2003) or that cognitive processes triggered by CF are similar across instructional settings. As Mitchell and Myles (2004) argued, "the underlying learning processes are essentially the same for more local and for more remote target languages, despite differing learning purposes and circumstances" (p. 6). We suggest that more qualitative types of research have the potential to contribute to "a better understanding of the relevant contextual variables that influence classroom learners' attentional biases toward one type of interactional feedback over another" (Lyster & Mori, 2006, p. 294).

Length of Instruction

Comparisons of how differential amounts of instruction impact CF effectiveness yielded one significant difference—namely, the between-group difference between short-to-medium (d = 0.57) and long treatments (d = 1.13). As is the case with other variables, these results point to the need for more research to further probe this topic; to our knowledge, there

has not been any research specifically focusing on the length of instruction as a variable that influences CF effectiveness. However, it might be the case that length of instruction alone is not a reliable index for SLA processes, as suggested by Norris and Ortega's (2000) acknowledgement that the small differences they observed between shorter and longer term treatments were likely due to a number of study variables, such as "the interaction of length and intensity of instruction with target structures, the interaction between treatment and type of outcome measures, and other moderator variables" (p. 487).

Learner Age

A simple regression analysis showed that the effects of CF are linearly related to learners' age, with the age factor accounting for 26.2% of the variance in the effects of CF for between-group contrasts and 16.4% of the variance in the effects of CF for within-group contrasts. This negative linear relationship between effects of CF and learners' age indicates that the effects of CF could be influenced by how old L2 learners are: The younger the learners are, the more they benefit from CF. It is possible that younger learners are especially sensitive to the impact of CF (Mackey & Oliver, 2002; Oliver, 2000) because it engages implicit learning mechanisms that are more characteristic of younger than older learners.

It may be the case that the advantage of younger learners is attributable to the fact that they received relatively longer instructional treatments (e.g., 7.5 hr in Ammar & Spada, 2006; 9 hr in Lyster, 2004) in comparison with adult learners (e.g., 90 min in Ellis, 2007; 90 min in Sheen, 2007). Although we acknowledge the possibility of various interpretations of the results, we speculate that the benefits that child and younger learners showed in our meta-analysis can be explained by age factors rather than by length of instruction, for the following reasons: Whereas only inconclusive results have been yielded by previous studies that investigated the impact of variable treatment length on L2 learning outcomes (e.g., Norris & Ortega, 2000), other SLA studies have demonstrated more conclusively that learners' age influences L2 developmental processes in significant ways (e.g., DeKeyser & Larson-Hall, 2005; Hyltenstam & Abrahamsson, 2003; but see also Birdsong, 2006; Marinova-Todd, Marshall, & Snow, 2000).

CONCLUSION

To measure the pedagogical effectiveness of CF in L2 classrooms, we performed a meta-analysis of quasi-experimental studies conducted

in classroom settings, focusing on various independent and dependent factors (i.e., timing of posttests, CF types, and types of outcome measures) and examining the effects of CF in relation to instructional settings, length of treatment, and age. Results showed that, irrespective of instructional settings, CF is facilitative of L2 development and that its impact is sustained at least until delayed posttests. Effect sizes in both between- and within-group contrasts were large for prompts, medium for recasts, and varied from large to medium for explicit correction. In comparison to the control groups, the prompt groups made significantly more progress than the recast groups between pretests and posttests. With respect to age, younger learners appeared especially sensitive to CF, benefiting from it significantly more than older learners. Additionally, students displayed different magnitudes of improvement through various outcome measurements, with the effects of CF most apparent in measures that elicit free constructed responses. The effects of long treatments proved to be significantly larger than those of short-to-medium treatments but not distinguishable from brief treatments, thus preventing any conclusions about length of instruction, especially in the absence of research that specifically investigates the relationship between length of treatments and CF effectiveness; the effects of CF might be more influenced by other variables such as CF intensiveness and types of outcome measures.

We conclude with some suggestions for further research as well as for practitioners in L2 classrooms. First and foremost, it is effective to employ CF in response to students' nontargetlike production because it contributes to target language development over time. That the effects of oral CF are durable and more apparent in free constructed-response measures than other types of measures points to the important role of CF as an effective form-focused instructional technique propitious for strengthening form-meaning connections and thus worthy of further exploration by teachers and researchers alike. In terms of CF types, the effects of prompts are larger than those of recasts in classroom settings, but a more substantial number of classroom studies that compare these CF types are necessary before drawing firm conclusions. Given the wide range of CF types that constitute both explicit correction and prompts, further research is also warranted to identify the components of these CF types that might contribute to their effectiveness. Additionally, further research is needed to probe the many learner characteristics that were not accounted for in this meta-analysis but that are known to mediate the effects of CF for individual learners (e.g., learners' proficiency, literacy levels, degree of anxiety, L1 background). As for age, the overall impact of CF in classroom settings might be greater for younger learners than for older learners; one suggestion provided here is that, especially for younger learners, L2 instruction with appropriate CF techniques

might have the potential to affect the development of target language accuracy. In other words, teachers might consider enhancing the pedagogical potential of CF not only by reformulating child learners' nontargetlike forms (i.e., recasting) but also by providing a variety of unambiguous signals and metalinguistic clues through various other types of CF.

Finally, given that the field of CF research has grown dramatically over the past 20 years, the time is ripe for research syntheses designed to move the research agenda forward. Yet, we stress the need to conduct meta-analyses in this domain through criteria of inclusion versus exclusion carefully selected in conjunction with research objectives, because theoretical and methodological constructs of CF studies are enormously diverse. Whereas our purpose was to explore the pedagogical value of CF by selecting only classroom-based CF studies, inclusion of the laboratory-based studies that we excluded from the current meta-analysis would also be of value to further investigate the effectiveness of CF in other interactional contexts.

NOTES

1. According to Cohen (1988), effect sizes are roughly classified as small ($0.20 \leq d < 0.50$), medium ($0.50 \leq d < 0.80$), or large ($0.80 \leq d$).

2. As is the case with previous meta-analyses, no attempt was made to contact the researchers to have further access to their data.

3. We used the pooled standard deviations rather than the standard deviations of the control group in the denominator, following Norris and Ortega's (2000) recommendation: "The standard deviation of any single group was considered particularly susceptible to sampling error, due to the small sample sizes of most studies within research domain" (p. 443).

4. Similarly, Norris and Ortega (2000) adopted the inclusive strategy and used multiple effect sizes generated by several of the unique sample studies "in order to provide the most representative picture of the instructional treatments that had received attention within the research domain" (p. 448).

5. In their analysis of specific versus general focus conditions, Russell and Spada (2006) found no significant differences, whereas Mackey and Goo (2007) found that, based on small samples, intensive CF is significantly more effective than extensive CF at the time of delayed (but not immediate) posttests. Future meta-analysis studies should include intensiveness of CF as an independent variable if a categorical distinction can be made between intensiveness and extensiveness.

6. Due to small sample sizes, only aggregate findings are shown rather than the outcome measures by instructional treatment (i.e., types of CF).

7. Because we did not find sufficient data for selected-response measures tasks ($n = 1$, $k = 8$), we do not make any argument about their relative effectiveness here.

8. Effect sizes for brief treatments were not further calculated for within-group contrasts because of the lack of sample studies ($n = 2$, $k = 4$).

9. In a further analysis of CF types, prompts, but not recasts, were found to be significantly related to age, which indicates that prompts might be more effective than recasts for younger learners, whereas prompts and recasts might be similarly effective for older learners. However, due to the small sample sizes in the statistical analysis of CF types in relation to age, we have not included the results here but plan on discussing these findings in another venue.

10. The symbol ** indicates studies that were included in the current meta-analysis (i.e., classroom-based studies), whereas the symbol * indicates studies that were not included in the current meta-analysis (i.e., lab-based studies).

REFERENCES[10]

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

Berns, M. (1990). 'Second' and 'foreign' in second language acquisition/foreign language learning: A sociolinguistic perspective. In B. VanPatten & J. F. Lee (Eds.), *Second language acquisition/foreign language learning* (pp. 3–11). Clevedon, UK: Multilingual Matters.

Birdsong, D. (2006). Age and second language acquisition and processing: A selective overview. *Language Learning*, *56*, 9–49.

Block, D. (2003). *The social turn in second language acquisition*. Washington, DC: Georgetown University Press.

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

*Carroll, S., Swain, M., & Roberge, Y. (1992). The role of feedback in adult second language acquisition: Error correction and morphological generalizations. *Applied Psycholinguistics*, *13*, 173–198.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). San Diego, CA: Academic Press.

**DeKeyser, R. (1993). The effect of error correction on L2 grammar knowledge and oral proficiency. *Modern Language Journal*, *77*, 501–514.

DeKeyser, R. (2003). Implicit and explicit learning. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 313–348). Oxford: Blackwell.

DeKeyser, R. (Ed.). (2007). *Practice in a second language: Perspectives from applied linguistics and cognitive psychology*. New York: Cambridge University Press.

DeKeyser, R., & Larson-Hall, J. (2005). What does the critical period really mean? In J. F. Kroll & A. M. B. De Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 89–108). Oxford: Oxford University Press.

Doughty, C. J. (2001). Cognitive underpinnings of focus on form. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 206–257). New York: Cambridge University Press.

Doughty, C. J. (2003). Instructed SLA: Constraints, compensation, and enhancement. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 257–310). Oxford: Blackwell.

Doughty, C. J., & Long, M. H. (Eds.). (2003). *The handbook of second language acquisition*. Oxford: Blackwell.

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

Doughty, C. J., & Williams, J. (Eds.). (1998). *Focus on forms in classroom second language acquisition*. New York: Cambridge University Press.

Ellis, R. (Ed.). (1999). *Learning a second language through interaction*. Amsterdam: Benjamins.

Ellis, R. (2001). Investigating form-focused instruction. *Language Learning*, *51*, 1–46.

Ellis, R. (2003). *Task-based language learning and teaching*. Oxford: Oxford University Press.

Ellis, R. (2006). Researching the effects of form-focussed instruction on L2 acquisition. *AILA Review*, *19*, 18–41.

**Ellis, R. (2007). The differential effects of corrective feedback on two grammatical structures. In A. Mackey (Ed.), *Conversational interaction in second language acquisition: A collection of empirical studies* (pp. 407–452). Oxford: Oxford University Press.

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

**Ellis, R., Loewen, S., & Erlam, R. (2006). Implicit and explicit corrective feedback and the acquisition of L2 grammar. *Studies in Second Language Acquisition*, *28*, 339–368.
**Ellis, R., Rosszell, H., & Takashima, H. (1994). Down the garden path: Another look at negative feedback. *JALT Journal*, *16*, 9–24.
Ellis, R., & Sheen, Y. (2006). Re-examining the role of recasts in L2 acquisition. *Studies in Second Language Acquisition*, *28*, 575–600.
Farrar, M. (1990). Discourse and the acquisition of grammatical morphemes. *Journal of Child Language*, *17*, 607–624.
Farrar, M. (1992). Negative evidence and grammatical morpheme acquisition. *Developmental Psychology*, *28*, 90–98.
Gass, S. M. (1997). *Input, interaction, and the second language learner.* Mahwah, NJ: Erlbaum.
*Han, Z. (2002). A study of the impact of recasts on tense consistency in L2 output. *TESOL Quarterly*, *36*, 543–572.
Havranek, G., & Cesnik, H. (2001). Factors affecting the success of corrective feedback. *EUROSLA Yearbook*, *1*, 99–122.
**Herron, C. (1991). The garden path correction strategy in the foreign language classroom. *The French Review*, *64*, 966–977.
**Herron, C., & Tomasello, M. (1988). Learning grammatical structures in a foreign language: Modeling versus feedback. *The French Review*, *61*, 910–922.
Hyltenstam, K., & Abrahamsson, N. (2003). Maturational constraints in SLA. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 539–588). Oxford: Blackwell.
*Inagaki, S., & Long, M. H. (1999). Implicit negative feedback. In K. Kanno (Ed.), *The acquisition of Japanese as a second language* (pp. 9–30). Amsterdam: Benjamins.
*Ishida, M. (2004). Effects of recasts on the acquisition of the aspectual form - *te i-(ru)* by learners of Japanese as a foreign language. *Language Learning*, *54*, 311–394.
*Iwashita, N. (2003). Positive and negative input in task-based interaction: Differential affects on L2 development. *Studies in Second Language Acquisition*, *25*, 1–36.
*Kim, H., & Mathes, G. (2001). Explicit vs. implicit corrective feedback. *The Korean TESOL Journal*, *4*, 57–72.
Krashen, S. (1981). *Second language acquisition and second language learning.* Oxford: Pergamon.
*Leeman, J. (2003). Recasts and second language development: Beyond negative evidence. *Studies in Second Language Acquisition*, *25*, 37–63.
Lightbown, P. M. (2008). Transfer appropriate processing as a model for class second language acquisition. In Z. Han (Ed.), *Understanding second language process* (pp. 27–44). Clevedon, UK: Multilingual Matters.
Lipsey, M., & Wilson, D. (2001). *Practical meta-analysis.* Thousand Oaks, CA: Sage.
Loewen, S. (2005). Incidental focus on form and second language learning. *Studies in Second Language Acquisition*, *27*, 361–386.
*Loewen, S., & Erlam, R. (2006). Corrective feedback in the chatroom: An experimental study. *Computer Assisted Language Learning*, *19*, 1–14.
**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. H. (1996). The role of the linguistic environment in second language acquisition. In W. C. Ritchie & T. K. Bhatia (Eds.), *Handbook of language acquisition: Vol. 2. Second language acquisition* (pp. 413–468). San Diego, CA: Academic Press.
Long, M. H. (2007). *Problems in SLA.* Mahwah, NJ: Erlbaum.
*Long, M. H., Inagaki, S., & Ortega, L. (1998). The role of negative feedback in SLA: Models and recasts in Japanese and Spanish. *Modern Language Journal*, *82*, 357–371.
Long, M. H., & Robinson, P. (1998). Focus on form: Theory, research, and practice. In C. J. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 15–41). New York: Cambridge University Press.
Lyster, R. (1998). Recasts, repetition, and ambiguity in L2 classroom discourse. *Studies in Second Language Acquisition*, *20*, 51–81.

Lyster, R. (2002). Negotiation in immersion teacher-student interaction. *International Journal of Educational Research*, *37*, 237–253.
**Lyster, R. (2004). Differential effects of prompts and recasts in form-focused instruction. *Studies in Second Language Acquisition*, *26*, 399–432.
Lyster, R. (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. *Studies in Second Language Acquisition*, *19*, 37–66.
**Mackey, A. (2006). Feedback, noticing, and second language development: An empirical study of L2 classroom interaction. *Applied Linguistics*, *27*, 405–430.
Mackey, A. (Ed.). (2007). *Conversational interaction in SLA: A collection of empirical studies*. Oxford: Oxford University Press.
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., & Oliver, R. (2002). Interactional feedback and children's L2 development. *System*, *30*, 459–477.
*Mackey, A., & Philp, J. (1998). Conversational interaction and second language development: Recasts, responses and red herrings? *Modern Language Journal*, *82*, 338–356.
Marinova-Todd, S., Marshall, B., & Snow, C. (2000). Three misconceptions about age and L2 learning. *TESOL Quarterly*, *34*, 9–34.
*McDonough, K. (2005). Identifying the impact of negative feedback and learners' response on ESL question development. *Studies in Second Language Acquisition*, *27*, 79–103.
*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.
McHoul, A. (1990). The organization of repair in classroom talk. *Language in Society*, *19*, 349–377.
Mitchell, R., & Myles, F. (2004). *Second language learning theories* (2nd ed.). London: Arnold.
**Muranoi, H. (2000). Focus on form through interaction enhancement: Integrating formal instruction into a communicative task in EFL classrooms. *Language Learning*, *50*, 617–673.
Nicholas, H., Lightbown, P. M., & 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.
Norris, J., & Ortega, L. (2006). The value and practice of research synthesis for language learning and teaching. In J. Norris & L. Ortega (Eds.), *Synthesizing research on language learning and teaching* (pp. 3–52). Amsterdam: Benjamins.
Oliver, R. (2000). Age differences in negotiation and feedback in classroom and pair work. *Language Learning*, *50*, 119–151.
*O'Relly, L. V., Flaitz, J., & Kromrey, J. D. (2001). Two modes of correcting communicative tasks: Recent findings. *Foreign Language Annals*, *34*, 246–257.
Ortega, L. (2009). *Understanding second language acquisition*. London: Hodder Education.
Philp, J., Mackey, A., & Oliver, R. (2008). Child's play? Second language acquisition and the younger learner in context. In J. Philp, R. Oliver, & A. Mackey (Eds.), *Second language acquisition and the younger learner: Child's play?* (pp. 3–23). Amsterdam: Benjamins.
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.

Russell, J., & Spada, N. (2006). The effectiveness of corrective feedback for second language acquisition: A meta-analysis of the research. In J. Norris & L. Ortega (Eds.), *Synthesizing research on language learning and teaching* (pp. 133–164). Amsterdam: Benjamins.

*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.

*Sauro, S. (2009). Computer-mediated corrective feedback and the development of L2 grammar. *Language Learning & Technology*, *13*, 96–120.

Segalowitz, N. (1997). Individual differences in second language acquisition. In A. M. B. De Groot & J. F. Kroll (Eds.), *Tutorials in bilingualism: Psycholinguistic perspectives* (pp. 85–112). Mahwah, NJ: Erlbaum.

Segalowitz, N. (2000). Automaticity and attentional skill in fluent performance. In H. Riggenbach (Ed.), *Perspectives on fluency* (pp. 200–219). Ann Arbor: University of Michigan Press.

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

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

**Sheen, Y. (2007). The effects of corrective feedback, language aptitude and learner attitudes on the acquisition of English articles. In A. Mackey (Ed.), *Conversational interaction in second language acquisition: A collection of empirical studies* (pp. 301–322). Oxford: Oxford University Press.

Spada, N. (1997). Form-focused instruction and second language acquisition: A review of classroom and laboratory research. *Language Teaching*, *29*, 73–87.

Spada, N., & Lightbown, P. M. (2009). Interaction in second/foreign language classrooms. In A. Mackey & C. Polio (Eds.), *Multiple perspectives on interaction: Second language research in honor of Susan M. Gass* (pp. 157–174). London: Routledge.

Spada, N., & Tomita, Y. (in press). Interactions between type of instruction and type of language feature: A meta-analysis. *Language Learning*.

Stern, H. H. (1983). *Fundamental concepts of language teaching*. Oxford: Oxford University Press.

**Takashima, H., & Ellis, R. (1999). Output enhancement and the acquisition of the past tense. In R. Ellis (Ed.), *Learning a second language through interaction* (pp. 173–188). Amsterdam: Benjamins.

**Tomasello, M., & Herron, C. (1989). Feedback for language transfer errors: The garden path technique. *Studies in Second Language Acquisition*, *11*, 385–395.

Truscott, J. (1999). What's wrong with oral grammar correction. *Canadian Modern Language Review*, *55*, 437–456.

APPENDIX

SUMMARY OF RESULTS BY STUDY

Types of CF	*d* Values (between-group)	*d* Values (within-group)
Herron and Tomasello (1988)[a]		
Explicit correction 1 (*n* = 16)		
Immediate	0.79	na
Explicit correction 2 (*n* = 16)		
Immediate	0.72	na
Tomasello and Herron (1989)		
Explicit correction (*n* = 16)		
Immediate	1.02	na
Delayed	0.77	na
Herron (1991)		
Explicit correction (*n* = 16)		
Immediate	1.63	na
Delayed	1.42	na
DeKeyser (1993)		
Explicit correction (*n* = 16)		
Immediate	0.40	0.07
Ellis et al. (1994)		
Explicit correction (*n* = 28)		
Immediate	0.38	0.97
Takashima and Ellis (1999)[b]		
Prompts		
Immediate	0.71	na
Delayed 1	0.72	na
Delayed 2	0.94	na
Muranoi (2000)[c]		
Recasts 1 (*n* = 30)		
Immediate	1.37	1.85
Delayed	0.96	1.59
Recasts 2 (*n* = 30)		
Immediate	0.38	0.61
Delayed	0.41	0.85
Lyster (2004)		
Recasts (*n* = 49)		
Immediate	0.94	0.83
Delayed	1.02	0.84
Prompts (*n* = 38)		
Immediate	1.58	1.58
Delayed	1.66	1.60

Continued

Appendix (Continued)

Types of CF	*d* Values (between-group)	*d* Values (within-group)
	Mackey (2006)[d]	
Recasts and prompts (*n* = 15)		
Immediate	0.16	na
	Ammar and Spada (2006)	
Recasts (*n* = 22)		
Immediate	0.67	0.81
Delayed	0.94	0.91
Prompts (*n* = 22)		
Immediate	1.14	1.50
Delayed	1.83	1.82
	Ellis et al. (2006)[e]	
Recasts (*n* = 12)		
Immediate	0.06	0.13
Delayed	0.42	0.44
Prompts (*n* = 12)		
Immediate	0.43	0.52
Delayed	0.63	0.79
	Ellis (2007)[e]	
Recasts (*n* = 12)		
Immediate	1.19	1.21
Delayed	0.62	1.12
Prompts (*n* = 12)		
Immediate	0.54	0.94
Delayed	0.76	1.45
	Loewen and Nabei (2007)[f]	
Recasts (*n* = 10)		
Immediate	0.15	0.70
Prompts 1 (*n* = 7)		
Immediate	0.26	0.60
Prompts 2 (*n* = 8)		
Immediate	–0.16	0.61
	Sheen (2007)	
Recasts (*n* = 26)		
Immediate	0.04	0.34
Delayed	0.15	0.42
Explicit correction (*n* = 26)		
Immediate	0.49	0.64
Delayed	0.63	0.71
	Yang and Lyster (this issue)	
Recasts (*n* = 25)		
Immediate	0.35	0.86
Delayed	0.32	0.56

Continued

Appendix (Continued)

Types of CF	*d* Values (between-group)	*d* Values (within-group)
Prompts (*n* = 22)		
Immediate	0.61	1.23
Delayed	0.73	1.09

Notes. The abbreviation "na" corresponds to not applicable.
[a]Explicit correction was directed to both feedback groups; one group (explicit correction 1) received feedback on direct object pronoun replacement and the other group (explicit correction 2) on negation.
[b]In Takashima and Ellis (1999), delayed posttests took place 2 weeks (delayed 1) and 7 weeks (delayed 2) after the intervention.
[c]In Muranoi (2000), the two recast groups are different in that one group (recasts 1) received explicit instruction on the target form but the other group (recasts 2) received comments on the content of their task performance.
[d]In Mackey's (2006) experiment, the distinction between recasts and clarification requests was labeled as recasts and negotiation rather than recasts and prompts.
[e]The intention of both studies (Ellis, 2007; Ellis et al., 2006) was to compare explicit (i.e., metalinguistic) and implicit (i.e., recasts) feedback. Ellis et al. operationalized explicit feedback as metalinguistic feedback, citing Lyster and Ranta's (1997) definition of this CF type: "comments, information, or questions related to the well-formedness of the learner's utterance" (p. 47). Therefore, we categorized Ellis's explicit feedback as prompts.
[f]In Loewen and Nabei's (2007) study, one prompt group (prompts 1) received clarification requests, whereas the other (prompts 2) received metalinguistic clues.