

# IMPLICIT AND EXPLICIT CORRECTIVE FEEDBACK AND THE ACQUISITION OF L2 GRAMMAR

Rod Ellis, Shawn Loewen, and Rosemary Erlam

*University of Auckland*

---

This article reviews previous studies of the effects of implicit and explicit corrective feedback on SLA, pointing out a number of methodological problems. It then reports on a new study of the effects of these two types of corrective feedback on the acquisition of past tense *-ed*. In an experimental design (two experimental groups and a control group), low-intermediate learners of second language English completed two communicative tasks during which they received either recasts (implicit feedback) or metalinguistic explanation (explicit feedback) in response to any utterance that contained an error in the target structure. Acquisition was measured by means of an oral imitation test (designed to measure implicit knowledge) and both an untimed grammaticality judgment test and a metalinguistic knowledge test (both designed to measure explicit knowledge). The tests were administered prior to the instruction, 1 day after the instruction, and again 2 weeks later. Statistical comparisons of the learners' performance on the posttests showed a clear advantage for explicit feedback over implicit feedback for both the delayed imitation and grammaticality judgment posttests. Thus, the results indicate that metalinguistic explanation benefited implicit as well as explicit knowledge and point to the importance of including measures of both types of knowledge in experimental studies.

---

This research was funded by a Marsden Fund grant awarded by the Royal Society of Arts of New Zealand. Researchers other than the authors who contributed to the research were Jenefer Philip, Satomi Mizutami, Keiko Sakui, and Thomas Delaney. Thanks go to the editors of this special issue and to two anonymous *SSLA* reviewers of a draft of the article for their constructive comments.

Address correspondence to: Rod Ellis, Department of Applied Language Studies and Linguistics, University of Auckland, Private Bag 92019, Auckland, New Zealand; e-mail: r.ellis@auckland.ac.nz.

Corrective feedback takes the form of responses to learner utterances that contain an error. The responses can consist of (a) an indication that an error has been committed, (b) provision of the correct target language form, or (c) metalinguistic information about the nature of the error, or any combination of these.

There has been a growing interest in the role of corrective feedback in SLA in the last decade. A number of descriptive studies based on data collected in classrooms (e.g., Panova & Lyster, 2002; Sheen, 2004) and on data collected in a laboratory-type setting (e.g., Iwashita, 2003; Mackey, Oliver, & Leeman, 2003; Philp, 2003) have examined the types of corrective feedback received by learners and the extent to which this feedback is noticed, or uptaken, or both by the learners. Experimental studies have attempted to examine the contribution that corrective feedback makes to acquisition (e.g., Ayoun, 2004; Han, 2002; Leeman, 2003; Lyster, 2004). This research has addressed, among other issues, the relative efficacy of implicit and explicit types of corrective feedback.

## THEORETICAL ISSUES

A distinction can be drawn between implicit/explicit *learning* and implicit/explicit *knowledge*. In the case of learning, the term *implicit* refers to “acquisition of knowledge about the underlying structure of a complex stimulus environment by a process that takes place naturally, simply and without conscious operations,” whereas *explicit learning* is “a more conscious operation where the individual makes and tests hypotheses in search for a structure” (N. Ellis, 1994, p. 2).

In the case of knowledge, the term *implicit* refers to knowledge that learners are only intuitively aware of and that is easily accessible through automatic processing, whereas *explicit knowledge* consists of knowledge that learners are consciously aware of and that is typically only available through controlled processing. Explicit knowledge might be linked to metalinguistic labels. These two types of knowledge are not mutually exclusive; that is, speakers can hold implicit and explicit representations of the same linguistic feature, as, for example, in the case of linguists who formulate explicit rules on the basis of their implicit knowledge of a language.

Schmidt (1994) stated that “implicit and explicit *learning* and implicit and explicit *knowledge* are related but distinct concepts that need to be separated” (p. 20). It can be argued, for example, that implicit knowledge is not entirely dependent on implicit learning but can arise as a product of learners intentionally practicing linguistic forms that they initially know explicitly (DeKeyser, 2003). It has also been argued that the development of implicit learning involves at least some degree of consciousness, as when learners notice specific linguistic forms in the input (Schmidt).

Corrective feedback differs in terms of how implicit or explicit it is. In the case of implicit feedback, there is no overt indicator that an error has been

committed, whereas in explicit feedback types, there is. Implicit feedback often takes the form of *recasts*, defined by Long (in press) as

a reformulation of all or part of a learner's immediately preceding utterance in which one or more non-target like (lexical, grammatical etc.) items are replaced by the corresponding target language form(s), and where, throughout the exchange, the focus of the interlocutors is on *meaning* not language as an object. (p. 2)

Recasts, therefore, provide positive evidence, but, as Nicholas, Lightbown, and Spada (2001) noted, it is not clear whether they provide negative evidence, as learners might have no conscious awareness that the recast is intended to be corrective. Explicit feedback can take two forms: (a) *explicit correction*, in which the response clearly indicates that what the learner said was incorrect (e.g., "No, not *goed—went*") and thus affords both positive and negative evidence or (b) *metalinguistic feedback*, defined by Lyster and Ranta (1997) as "comments, information, or questions related to the well-formedness of the learner's utterance" (p. 47)—for example, "You need past tense," which affords only negative evidence.

It can also be argued that recasts and explicit corrective strategies differ in terms of whether they cater to implicit or explicit learning. For Long (1996), recasts work for acquisition precisely because they are implicit, connecting linguistic form to meaning in discourse contexts that promote the microprocessing (i.e., noticing or rehearsing in short-term memory) required for implicit language learning. Doughty (2001), building on Long's rationale for focus-on-form, argued that recasts constitute the ideal means of achieving an "immediately contingent focus on form" and afford a "cognitive window" (p. 252) in which learners can rehearse what they have heard and access material from their interlanguage. In contrast, explicit corrective feedback strategies, such as metalinguistic feedback, are more likely to impede the natural flow of communication and to activate the kind of learning mechanisms that result in explicit rather than implicit second language (L2) knowledge. However, such a view is not unproblematic.

First, it is not certain that all recasts are as implicit as Long (1996) and Doughty (2001) assumed. Some recasts are quite explicitly corrective. Indeed, the kind of corrective recasts that Doughty and Varela (1998) employed in their experimental study were remarkably explicit. They were preceded by a repetition of the learner's utterance with the erroneous elements highlighted by emphatic stress. If the learner did not self-correct, recasts with emphatic stress to draw attention to the reformulated elements followed. Thus, if the corrective force of the recast becomes self-evident, it is difficult to argue that it constitutes an implicit or even a relatively implicit technique. Second, recasts can only work for acquisition if learners notice the changes that have been made to their own utterances, and there are reasons to believe that they do not always do so. Lyster (1998) has shown that the levels of repair in uptake following recasts are notably lower than those following more explicit types of

feedback. The findings from Lyster's research, which examined immersion classrooms in Canada, were corroborated by Sheen (2004), who found that repair occurred less frequently following recasts than following explicit correction and metalinguistic feedback in four different instructional contexts (immersion, Canadian English as a second language [ESL], New Zealand ESL, and Korean English as a foreign language). Even though repair cannot be taken as a measure of learning, it is reasonable to assume that it constitutes a measure of whether learners have noticed the key linguistic forms (although noticing can occur even if there is no uptake). Further evidence of the difficulty that learners might experience in attending to the key forms comes from a study by Mackey, Gass, and McDonough (2000), which demonstrated that learners often failed to perceive recasts that contained morphosyntactic reformulations as corrections. Finally, we cannot be certain that recasts promote acquisition of implicit knowledge. Indeed, it is entirely possible that recasts result in explicit knowledge, as demonstrated in Long, Inagaki, and Ortega (1998); in this study, all eight students who had learned the target structure (Spanish adverb word order) through recasts were able to explicitly and correctly formulate an explanation of the rule. Thus, there are some doubts as to how effective recasts are in promoting learning as well as to what kind of learning and knowledge they cater.

A case can also be made for the contribution of corrective strategies that are self-evidently corrective to learning. Carroll's (2001) autonomous induction theory posits that feedback can only work for acquisition if the corrective intentions of the feedback are recognized by the learner. Additionally, learners must be able to locate the error; Carroll noted that "most of the indirect forms of feedback do not locate the error" (p. 355). Recasts do not overtly signal that an error has been made and might assist in locating the error depending on whether the recast is full (i.e., the whole erroneous utterance is reformulated) or partial (i.e., only the erroneous part of the utterance is reformulated), as Sheen's (in press) study indicates. In contrast, explicit types of feedback not only make the corrective force clear to the learner but also give clues as to the exact location of the error. As such, they might be more likely to induce learners to carry out the cognitive comparison between their error and the target form (R. Ellis, 1994), which is believed to foster acquisition (Schmidt, 1994).

Connectionist models also lend support to explicit error correction. N. C. Ellis (2005) distinguished between the mechanisms of conscious and unconscious learning, emphasizing the role of attention and consciousness in the former and of connectionist learning in the latter. He proposed the learning sequence in (1).

- (1) external scaffolded attention → internally motivated attention → explicit learning → explicit memory → implicit learning → implicit memory, automatization, and abstraction

Ellis went on to suggest that "conscious and unconscious processes are dynamically involved together in every cognitive task and in every learning epi-

sode" (p. 340). Although he did not suggest that explicit corrective feedback is the ideal mechanism for achieving this continuous synergy (indeed, his discussion of feedback is restricted to recasts), it would seem that the metalinguistic time-outs from communicating afforded by explicit correction constitute a perfect context for melding the conscious and unconscious processes involved in learning. Within the context of a single interactional exchange, such a time-out creates an opportunity for learners to traverse the learning sequence sketched out in (1). Of course, no single exchange can guarantee that the targeted form will enter implicit memory, but repeated exchanges—directed at the same linguistic form—might be expected to do so. Thus, according to such a theoretical perspective, explicit corrective feedback caters not just to explicit learning and explicit memory but also to implicit learning and implicit memory.

## PREVIOUS RESEARCH ON CORRECTIVE FEEDBACK

This review will focus on studies that have compared the effects of implicit and explicit corrective feedback on L2 acquisition and will address, in particular, methodological issues in keeping with the theme of this special issue.<sup>1</sup> There have been a number of other studies that investigated separately whether either implicit or explicit corrective feedback facilitates acquisition. Nicholas et al. (2001), Long (in press), and Ellis and Sheen (in press) provided reviews of the research on recasts. In general, the recast studies demonstrated that implicit feedback of this kind can have a beneficial effect on acquisition, especially when the recasts are more explicit in nature (as in Doughty & Varela, 1998). Other studies demonstrated that explicit feedback is of value. Carroll, Roberge, and Swain (1992), for example, found that a group that received explicit corrective feedback directed at two complex French noun suffixes (*-age* and *-ment*) outperformed a group that received no feedback, although no generalization of learning to nouns not presented during the treatment occurred. Thus, the recast and explicit feedback studies demonstrated that both types of feedback can be effective.

Table 1 summarizes 11 studies that have compared implicit and explicit corrective feedback. It is not easy to come to clear conclusions about what these studies reveal due to a number of factors. First, whereas some of the studies are experimental in nature (e.g., Carroll, 2001; Carroll & Swain, 1993; Lyster, 2004; Rosa & Leow, 2004), others are not (e.g., DeKeyser, 1993; Havranek & Cesnik, 2003), as this second group of researchers investigated corrective feedback through post hoc analyses of normal classroom lessons. Second, the studies vary in terms of whether they involved laboratory, classroom, or computer-based interaction. Third, the nature of the treatment activities performed by the learners in these studies differed considerably. In some cases, the activities involved fairly mechanical exercises (e.g., Carroll; Carroll & Swain; Nagata, 1993), in others they involved communicative tasks (e.g., Leeman, 2003; Muranoi, 2000; Rosa & Leow), and in others they involved a mixture of the two (DeKeyser). Fourth, the treatment also differed in terms of whether it

**Table 1.** Studies comparing the effects of different types of corrective feedback

Study	Participants	Target structure	Design	Tests	Results
Carroll & Swain (1993)	100 Spanish adult ESL learners (low intermediate)	Dative verbs	Groups: (A) direct metalinguistic feedback, (B) explicit rejection, (C) recast, (D) indirect metalinguistic feedback, (E) control.  Treatment: two feedback sessions, each followed by recall (i.e., production without feedback).	Recall production tasks following each feedback session.	All of the treatment groups performed better than the control group on both recall tasks. Group A outperformed the other groups.
Nagata (1993)	32 second-year university learners of L2 Japanese	Japanese passive structures; verbal predicates and particles	Groups: (A) received feedback indicating what was missing or not expected, (B) received same feedback + metalinguistic explanations.  Treatment: Learners performed computer-based exercises requiring them to respond to sentences produced by an imaginary partner. Sentences were computer-parsed and feedback on errors provided on the basis of group membership.	Written test using same format as treatment task.	Group B significantly outperformed group A on particles but not verbal predicates. Learners expressed preference for metalinguistic explanation.
DeKeyser (1993)	25 Dutch high school seniors learning L2 French	Variety of features, predominantly morphosyntactic	Groups: (A) extensive explicit corrective feedback during normal class activities, (B) limited explicit corrective feedback.  Treatment: 10 class periods.	Three oral communication tasks (interview, picture description, and story-telling).  Fill-in-the-blank test.  Tests administered twice.	No statistically significant differences evident between groups A and B. Learners with high previous achievement, high language aptitude, high extrinsic motivation, and low anxiety benefited the most from error correction.

Muranoi (2000)	114 first-year Japanese college students	Indefinite article to denote new information	Groups: (A) interaction enhancement (IE) by means of requests for repetition and recasts in communicative task + formal debriefing (explicit grammar explanation), (B) IE + meaning-focused debriefing, (C) control (no IE with meaning-focused debriefing).	Grammaticality judgment test. Oral production task. Written production task. Two posttests 5 weeks apart.	Both experimental groups outperformed the control group on both posttests. Group A outperformed group B on posttest 1 but not on posttest 2.
Kim & Mathes (2001)	20 Korean adult ESL learners (high beginners and intermediate)	Dative verbs	Groups: (A) received explicit metalinguistic feedback, (B) received recasts. Treatment period: Feedback was presented in two sessions 1 week apart each followed by production with no feedback.	Controlled production tasks (as in the treatment) without feedback.	Differences between performance on first and second production tasks were not significant. Differences between groups for gains in production were not significant. Learners expressed preference for explicit feedback.
Carroll (2001)	100 adult low-intermediate ESL learners	Forming nouns from verbs (e.g., "help" (V) → "help/helping" (N) and distinguishing <i>thing</i> and <i>event</i> nouns)	Groups: as in Carroll & Swain (1993).	Elicited verb-noun conversions in a sentence format.	All types of feedback helped learners to learn the items targeted by the feedback but only explicit metalinguistic information (group A) and indirect prompting (group D) enabled learners to form a generalization. Recasts did not facilitate generalization.
Havranek & Cesnik (2003)	207 university students specializing in English	Variety of English phonological, lexical, and grammatical features	Data on 1700 corrective feedback episodes from normal English lessons.	Class-specific tests (translation, correction, reading aloud, and written and spoken completion tasks) directed at corrected items.	Effectiveness of corrective feedback techniques was, in order: (1) elicited self-correction, (2) explicit rejection + recast, (3) recast alone.

(continued)

**Table 1** (Continued)

Study	Participants	Target structure	Design	Tests	Results
Leeman (2003)	74 first-year university learners of Spanish	Spanish noun-adjective agreement	<p>Groups: (A) recast group, (B) negative evidence group (source or problem indicated but not corrected), (C) enhanced salience with no feedback, (D) control group.</p> <p>Treatment: Participants performed communicative task one-on-one with researcher.</p>	Post- and delayed post-picture description tasks.	Only groups A and C outperformed the control group on any posttest measure. No difference between A and C.
Sanz (2003)	28 first-year university learners of Spanish	Position of clitic pronouns between object and verb	<p>Groups: (A) explicit metalinguistic feedback, (B) implicit feedback (e.g., "Sorry, try again").</p> <p>Treatment: Computer-delivered input processing instruction without prior explicit instruction.</p>	<p>Interpretation tests.</p> <p>Production tests: (a) sentence completion and (b) written video retelling.</p>	Both groups significantly increased ability to interpret and accurately produce the target structure with no difference between the groups on any measure.

Lyster (2004)	148 (grade 5) 10–11-year-olds in a French immersion program	French grammatical gender (articles + nouns)	Groups: (A) received form-focused instruction (FFI) + recasts, (B) FFI + prompts (including explicit feedback), (C) FFI only, (D) control group.	Binary choice test. Text completion test (oral production tasks). Object identification test. Picture description test. Two posttests: The second administered 8 weeks after the first.	Group B was only group to outperform control group on all 8 measures (posttests 1 and 2). Group A outperformed control group on 5 of 8 measures. Group C outperformed control group on 4 of 8 measures. Statistically significant differences were found between groups B and C but not between A and B.
Rosa & Leow (2004) <sup>2</sup>	100 adult university learners of L2 Spanish enrolled in advanced courses	Contrary to the fact conditional sentences in the past	Groups: (A) explicit feedback to both correct and incorrect responses involving metalinguistic explanation + opportunity to try again if incorrect, (B) implicit feedback indicating whether the answer was right or wrong, (C) control group.  Treatment: Computer-based exposure to input-based jigsaw task characterized by task essentialness.	Three multiple-choice recognition tests and three written controlled production tests.  Immediate and delayed posttests.	Results presented in terms of old and new items. Recognition tests: A statistically significant difference evident between groups A and B for new but not old items.  Production tests: A statistically significant difference was evident for the old but not the new items. Both groups outperformed the control group.

involved output processing (the vast majority of the studies) or input processing (Rosa & Leow; Sanz, 2003). Fifth, the studies vary considerably in how they operationalized implicit and explicit feedback. Given the importance of this variable, it is discussed in greater detail later in this section. Sixth, variation is evident in how learning was measured: Some studies utilized metalinguistic judgments (e.g., Muranoi), selected response, or constrained constructed response formats (e.g., Havranek & Cesnik; Rosa & Leow), all of which might be considered to favor the application of explicit knowledge, whereas others opted for a free constructed response format (e.g., Leeman), which is more likely to tap implicit knowledge. Finally, the studies differ in another important respect: Some included an explicit explanation of the grammatical target prior to the practice activity (e.g., Lyster; Muranoi), whereas others did not (e.g., Leeman; Sanz). These differences in design reflect the different purposes of the studies, not all of which were expressly intended to compare implicit and explicit corrective feedback.

Implicit feedback in these studies has typically taken the form of recasts (Carroll, 2001; Carroll & Swain, 1993; Kim & Mathes, 2001; Leeman, 2003; Lyster, 2004). However, Muranoi (2000) employed both recasts and requests for repetition. Sanz (2003) made use of only requests for repetition ("Sorry, try again."). In Havranek and Cesnik's (2003) classroom study, which investigated naturally occurring corrective feedback, a variety of more or less implicit forms were identified, including recast, rejection + recast, and recast + repetition. This bears out the claims of Nicholas et al. (2001) and Ellis and Sheen (in press) that recasts actually vary considerably in how implicit or explicit they are. It should be noted, therefore, that the recasts used in the different studies might not have been equivalent in their degree of implicitness versus explicitness.

Explicit feedback has also been operationalized in very different ways. A minimal form of explicit feedback consists of simply indicating that an error has been committed (e.g., Carroll & Swain's, 1993, *explicit rejection* or Leeman's, 2003, *negative evidence*). Rosa and Leow's (2004) *implicit condition* actually consisted of indicating whether the learners' answers were right or wrong and, thus, might have been more accurately labeled *semiexplicit*. Carroll (2001), Carroll and Swain, Nagata (1993), and DeKeyser (1993) distinguished between a form of minimal explicit feedback, involving some specification of the nature of the error, and extensive corrective feedback, involving more detailed metalinguistic knowledge. Lyster's (2004) *prompts* consisted of clarification requests, repetitions (with the error highlighted suprasegmentally), metalinguistic clues, and elicitation of the correct form. Therefore, in Lyster's sense, prompts include both implicit and explicit forms of feedback. The nonexperimental classroom studies (DeKeyser; Havranek & Cesnik, 2003) inevitably involved a variety of explicit forms of correction. All of these studies examined explicit correction provided online immediately following learner utterances that contained errors. In contrast, Muranoi's (2000) study investigated the effects of providing an explicit grammar explanation after the treatment task had been completed.

Given the substantial differences in the purposes and designs of these studies, care needs to be taken in any attempt to generalize the findings. However, overall, the results point to an advantage for explicit over implicit corrective feedback in studies in which the treatment involved production. Carroll and Swain (1993) and Carroll (2001) reported that the group that received direct metalinguistic feedback outperformed all of the other groups in the production of sentences involving dative verbs and noun formation and that this type of feedback aids generalization to novel items. Muranoi (2000) found that the group that received formal debriefing (which included metalinguistic information) outperformed the group that received meaning-focused debriefing, although only on the immediate posttest. Havranek and Cesnik (2003) found that bare recasts were the least effective form of feedback in their classroom study. Lyster (2004) reported that the group that received prompts (which included metalinguistic feedback) performed better than the group that received recasts on both immediate and delayed posttests. There is also some evidence that a comparison between two types of explicit feedback will show that the more detailed metalinguistic feedback works better (Nagata, 1993; Rosa & Leow, 2004). It is also worth noting that the two studies that asked learners about what type of feedback they preferred (Kim & Mathes, 2001; Nagata) reported a clear preference for more explicit feedback.

However, not all of the studies point to an advantage for explicit feedback. DeKeyser (1993) found no difference between the group that received extensive explicit feedback and the group that received limited explicit feedback. Nevertheless, his study indicated that when individual difference factors, such as the learners' proficiency and language aptitude, were taken into account, the more explicit feedback was of greater benefit to the more able learners. Kim and Mathes (2001), in a study that replicated that of Carroll and Swain (1993), also failed to find any statistically significant differences in the scores of the explicit and implicit groups. Explicit feedback that consists of simply indicating that a problem exists does not appear to be helpful (Leeman, 2003). In the one study that examined feedback as part of input-processing instruction (Sanz, 2003), explicit metalinguistic feedback did not confer any advantage. It is also important to recognize that these studies provide evidence that implicit methods of feedback can assist learning. The implicit groups in Carroll and Swain, Carroll (2001), Muranoi (2000), Leeman, and Lyster (2004) all scored higher than the control groups on the posttests.

The main limitation of the research to date lies in the method of testing. As noted previously, most of the studies did not include tests that can be considered valid measures of implicit knowledge (i.e., tests that call on learners to access their linguistic knowledge rapidly online). The kinds of test used (grammaticality judgment tests, sentence completion, picture prompt tests, translation tests) favored the use of explicit knowledge. It can be argued, therefore, that they were biased in favor of explicit corrective feedback. The studies that included a test likely to measure implicit knowledge did not provide clear comparisons of the effects of explicit and implicit feedback. For example, Muranoi

(2000) did not examine online feedback; in his study, feedback was provided after the treatment tasks were completed. Leeman (2003), as already pointed out, did not examine explicit feedback that contained metalinguistic explanations. Lyster (2004) did not examine metalinguistic clues separately from other types of nonexplicit feedback designed to elicit the negotiation of form.

The study reported in this article investigated the following research question:

Do learners learn more from one type of corrective feedback than from another type?

This study was designed to provide a precise comparison between implicit and explicit corrective feedback by operationalizing these constructs in terms of (a) partial recasts of those portions of learners' utterances that contained an error and (b) metalinguistic explanations in which the learner's error was repeated and followed by metalinguistic information about the target language rule but the correct target language form was not provided. The effects of the corrective feedback on learning were assessed by means of tests designed to measure learning of both implicit and explicit L2 knowledge.

## METHOD

The present study compares the effectiveness of two types of corrective feedback: explicit error correction in the form of metalinguistic information and implicit error correction in the form of recasts. Group 1 received implicit feedback (recast group), group 2 received explicit feedback (metalinguistic group), and group 3 (a testing control) had no opportunity to practice the target structure and, thus, received no feedback. The relative effectiveness of both types of feedback was assessed on an oral elicited imitation test, a grammaticality judgment test, and a test of metalinguistic knowledge. There were three testing times: a pretest, an immediate posttest, and a delayed posttest. The target grammatical structure was past tense *-ed*.

## Participants

The study was conducted in a private language school in New Zealand. Three classes of students ( $n = 34$ ) were involved. The school classified these classes as lower intermediate, according to scores on a placement or a previous class achievement test. Information obtained from a background questionnaire showed that the majority of learners (77%) were of East Asian origin. Most of them had spent less than a year in New Zealand (the mean length of stay was just over 6 months). The mean age of all participants was 25 years. The learners indicated that they had been formally engaged in studying English for any-

where from 8 months to 13 years, with an average length of time of 7 years. Around 44% of participants indicated that their studies had been mainly formal (grammar-oriented) in nature, whereas 30% had received mainly informal instruction, and the rest had received a mixture of both formal and informal instruction.

The teaching approach adopted by the school placed emphasis on developing communicative skills in English. Learners received between 3 and 5 hr of English language instruction a day, for which they were enrolled as part-time or full-time students. Classes were arbitrarily assigned to one of the two treatment options (group 1 = 12 students, group 2 = 12 students) or to the control group option (group 3 = 10 students).

### Target Structure

Regular past tense *-ed* was chosen as the target structure for two reasons. First, learners at the lower intermediate level are likely to already be familiar with and have explicit knowledge of this structure. Our purpose was not to examine whether corrective feedback assists the learning of a completely new structure, but whether it enables learners to gain greater control over a structure they have already partially mastered. Pretesting demonstrated that this was indeed the case: On the grammaticality judgment test, the learners scored a mean of 75% on past tense *-ed*. The second reason was that past tense *-ed* is known to be problematic for learners and to cause errors (e.g., Doughty & Varela, 1998); thus, it was hypothesized that although learners at this level would have explicit knowledge of this structure, they would make errors in its use, especially in a communicative context and especially in oral production (oral production poses a problem because of Asian learners' phonological difficulties in producing consonant clusters with final [t] or [d]). Once again, pretesting demonstrated that this was indeed the case: On the oral elicited imitation test, which served as a measure of unplanned language use, the learners scored a mean of 30% on regular past tense *-ed*.

Regular past tense *-ed* is typically introduced in elementary and lower intermediate textbooks, but it is not among the morphemes acquired early (Dulay & Burt, 1974; Makino, 1980). It is acquired after such morphemes as articles, progressive *-ing*, and plural *-s* but before such morphemes as long plural (*-es*) and third person *-s*. The typical error made by learners is the use of the simple or present form of the verb in place of *V-ed*: \*Yesterday I *visit* my sister. Hawkins (2001) noted that some L2 learners "have difficulty in establishing the regular pattern (for past tense) at all" (p. 65).

### Instructional Materials

For the purposes of the study, each experimental group received the same amount of instruction (i.e., a total of 1 hr over 2 consecutive days during which

they completed two different half-hour communicative tasks). The control group continued with their normal instruction. They did not complete the tasks and did not receive any feedback on past tense *-ed* errors.

The tasks were operationalized according to R. Ellis' (2003) definition of *tasks*; that is, they included a gap, they required learners to focus primarily on meaning and to make use of their own linguistic resources, and they had a clearly defined outcome. They constituted what Ellis called *focused tasks*; in other words, they were designed to encourage the use of particular linguistic forms and, to this end, learners were provided with certain linguistic prompts (see the description of each task).

**Task 1 (Day 1).** Learners were assigned to four triads. Each triad was given the same picture sequence, which narrated a short story. They were also given one of four versions of a written account of the same story. Each version differed in minor ways from the others. Learners were told that they would have only a couple of minutes to read the written account of the story and that they needed to read it carefully because they would be asked to retell it in as much detail as possible. They were not allowed to make any written notes. The stories were removed and replaced with the list of verbs in (2) that learners were told they would need in order to retell the story.

(2) visit live walk turn kill want follow attack laugh point stay watch

Learners were given about 5 min to plan the retelling of their story. They were told that they would not be able to use any prompts other than the picture sequence and verb list. The opening words of the story were written on the board, to clearly establish a context for past tense: "Yesterday, Joe and Bill. . ." Learners were then asked to listen to each triad's collective retelling of the story. They were also told that each triad had been given a slightly different version of the same story and that they were to listen carefully to identify what was different.

**Task 2 (Day 2).** Learners were once again assigned to triads. Each triad was given a picture sequence depicting a day in the life of one of two characters (Gavin or Peter). Each picture sequence was different. Pictures were chosen to depict actions that would require the use of verbs with regular past tense *-ed* forms. Learners were given 5 min to prepare for recounting the day of either Gavin or Peter. Again, they were not allowed to take any written notes.

Each triad was told to begin their account with "Yesterday Peter/Gavin had a day off." Learners in the other triads who were listening to the narrated story were provided with an empty grid and pictures that they were to place on the grid in the appropriate sequence, according to the narration. One picture card did not fit, and learners were told they would be asked to identify which card remained.

## Instructional Procedures

The same instructor—one of the researchers—was responsible for conducting both tasks. The learners had not met the instructor prior to the first treatment session. An observer sat in the classroom during each session to manually record on paper all instances of the use of the target structure and each instance of corrective feedback. The treatment sessions were also audio-recorded.

The learners received corrective feedback while they performed the tasks. Group 1 received implicit feedback in the form of recasts, as in (3).

- (3) Learner: . . . they saw and they follow follow follow him  
Researcher: Followed  
Learner: Followed him and attacked him.

The recasts were typically declarative and of the partial type and, as such, might be considered to lie at the explicit end of the implicit-explicit continuum for recasts (see Sheen, *in press*). Nevertheless, because they intruded minimally into the flow of the discourse, they might not have been very salient to the learners.

The learners in group 2 received explicit feedback in the form of metalinguistic information, as in (4).

- (4) Learner: He kiss her  
Researcher: Kiss—you need past tense.  
Learner: He kissed

In this example, which was typical of the corrective feedback episodes in this study, the instructor first repeated the error and then supplied the metalinguistic information. It is important to note that although corrective feedback was directed at individual learners, the task was designed to ensure that the attention of the whole class was focused as much as possible on the speaker at these times.

Table 2 indicates the number of target forms that were elicited during each task and the total number of incorrectly produced forms. The number of instances of feedback is also given.

## Testing Instruments and Procedures

Five days prior to the start of the instructional treatments, the learners involved in the study signed the consent forms, as required by the University of Auckland Human Participants Ethics Committee, and completed all of the pretests. The immediate posttesting was completed the day after the second (and last) day of instruction, and the delayed posttesting was completed 12 days later. During each testing session, three tests were administered in

**Table 2.** Number of target forms elicited and instances of feedback

Type of feedback	Total target forms elicited <sup>a</sup>	Total incorrect target forms elicited	Instances of feedback
Recast			
Lesson 1	44	32	24
Lesson 2	48	30	28
Metalinguistic			
Lesson 1	52	21	17
Lesson 2	50	24	23

<sup>a</sup>Correct and incorrect.

the following order: untimed grammaticality judgment test, metalinguistic knowledge test, oral imitation test. The oral imitation test (Erlam, in press) was intended to provide a measure of the learners' implicit knowledge, whereas the untimed grammaticality judgment test (especially the ungrammatical sentences in this test; see R. Ellis, 2005) and the metalinguistic test were designed to provide measures of learners' explicit knowledge.

R. Ellis (2004, 2005) discussed the theoretical grounds for these claims. He argued that tests of implicit knowledge need to elicit use of language where the learners operate by feel, are pressured to perform in real time, are focused on meaning, and have little need to draw on metalinguistic knowledge. In contrast, tests of explicit knowledge need to elicit a test performance in which the learners are encouraged to apply rules, are under no time pressure, are consciously focused on form, and have a need to apply metalinguistic knowledge. The oral imitation test was designed to satisfy the criteria for tests of implicit knowledge, whereas the untimed grammaticality judgment test and the metalinguistic test were designed to meet the criteria for tests of explicit knowledge. The tests are described in detail in the following subsections.

**Oral Imitation Test.** This test consisted of a set of 36 belief statements. Statements were grammatically correct ( $n = 18$ ) or incorrect ( $n = 18$ ). They consisted of 12 statements that targeted simple past tense *-ed*, 12 that targeted comparative adjectives (a focus of another study), and 12 distracter items. Examples of the past tense *-ed* items are given in (5).

- (5) a. Everyone liked the movie *Star Wars*.  
 b. \*An American invent Microsoft Word.

Statements included target items introduced during the instructional treatments (old items) and new items. The statements containing new items were designed to test whether learners were able to generalize what they had learned to new vocabulary items. Eight of the 12 statements targeting past tense *-ed* presented the target structure in the context of new items.

Each statement was presented orally, one at a time, on an audiotape: Test-takers were required to first indicate on an answer sheet whether they agreed with, disagreed with, or were not sure about the statement. They were then asked to repeat the statement orally in correct English. Pretest training presented learners with both grammatical and ungrammatical statements (not involving past tense *-ed*) to practice with, and they were given the correct responses to these items. Learners' responses to all items were audio-recorded. These were then analyzed to establish whether obligatory occasions for use of the target structure had been established. Errors in structures other than the target structure were not considered. Each imitated statement was allocated a score of either 1 (the grammatically correct target structure was correctly imitated or the grammatically incorrect target structure was corrected) or 0 (the target structure was avoided, the grammatically correct target structure was attempted but incorrectly imitated, or the grammatically incorrect target structure was imitated but not corrected). If a learner self-corrected, then only the initial incorrect production was scored, as it was felt that this would provide the better measure of learners' implicit knowledge. Scores were expressed as percentage correct. Three versions of the test were created for use over the three testing sessions; in each, the same statements were used but presented in a different order. Reliability (Cronbach's alpha) for the pretest was .779. For more information about the theoretical rationale for this test and its design, see Erlam (in press).

**Grammaticality Judgment Test.** This was a pen-and-paper test consisting of 45 sentences. Fifteen sentences targeted past tense *-ed*, and the remainder targeted 30 other structures. Of the 15 sentences, 7 were grammatically correct and 8 were grammatically incorrect. Sentences were randomly scrambled in different ways to create three versions of the test. Test-takers were required to (a) indicate whether each sentence was grammatically correct or incorrect, (b) indicate the degree of certainty of their judgment (as proposed by Sorace, 1996) by typing a score on a scale marked from 0% to 100% in the box provided, and (c) self-report whether they used a rule or feel for each sentence. Learners were given six sentences to practice on before beginning the test. Each item was presented on a new page, and test-takers were told that they were not allowed to turn back to look at any part of the test that they had already completed. For past tense *-ed*, 7 of the 15 statements presented the target structure in the context of new vocabulary and 8 presented the target structure in the context of vocabulary included in the instruction. Learners' responses were scored as either correct (1 point) or incorrect (0 points). In addition to the total score, separate scores for grammatical and ungrammatical test items and also for new and old verb items were calculated. Reliability (Cronbach's alpha) for the pretest was .63. Test-retest reliability (Pearson's  $r$ ) was calculated for the control group ( $n = 10$ ) only. For the pretest and immediate posttest, it was .65 ( $p < .05$ ), and for the pretest and delayed posttest, it was .74 ( $p < .05$ ).

**Metalinguistic Knowledge Test.** Learners were presented with five sentences and were told that they were ungrammatical. Two of the sentences contained errors in past tense *-ed*. The part of the sentence containing the error in each example was underlined. Learners were asked to (a) correct the error and (b) explain what was wrong with the sentence (in English, using their own words). They were shown two practice examples. As in the previous test, each item was presented on a new page and test-takers were told that they were not allowed to turn back. Learners were scored one point for correcting the error and one point for a correct explanation of the error. A percentage accuracy score was calculated.

## Analysis

Descriptive statistics for the three groups on all three tests were calculated. On the oral imitation test and the grammaticality judgment test, (a) total scores, (b) separate scores for grammatical and ungrammatical items, and (c) separate scores for old and new items were calculated. The decision to examine grammatical and ungrammatical items separately was motivated by previous research (see R. Ellis, 2005), which showed that these might measure different types of knowledge (i.e., ungrammatical sentences provide a stronger measure of explicit knowledge). The decision to examine old items (i.e., items that tested verbs included in the instructional treatment) versus new items (i.e., verbs not included in the instructional treatment) was motivated by the wish to examine whether the instruction resulted only in item learning or whether there was also evidence of system learning.

The *t*-tests showed that there were statistically significant differences among the groups on the oral imitation and grammaticality judgment pretests. To take account of this difference, analyses of covariance (ANCOVAs; with pretest scores as the covariate) were computed to investigate to what extent group differences on the two posttests were statistically significant.

## RESULTS

### Oral Imitation Test

The descriptive statistics for regular past tense on the imitation test (see Table 3) show a range in overall accuracy from 24% to 39% on the pretest. The scores increase on both posttests. The ungrammatical items have lower accuracy scores than the grammatical items.

The results of the ANCOVAs reveal that there is a significant difference between the groups on their pretest scores. Once these differences are taken into account, there is no effect for group on the immediate posttest,  $F(2, 34) = 0.961$ ,  $p = .394$ ; however, there is on the delayed posttest,  $F(2, 34) = 7.975$ ,  $p < .01$ . The post hoc contrasts for the delayed posttest show that the metalin-

**Table 3.** Imitation test results

Groups	Pretest		Immediate posttest		Delayed posttest	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Recast						
Grammatical	0.278	0.278	0.403	0.279	0.514	0.180
Ungrammatical	0.194	0.282	0.319	0.240	0.375	0.334
Total	0.236	0.164	0.361	0.228	0.444	0.223
Metalinguistic						
Grammatical	0.444	0.192	0.618	0.257	0.736	0.194
Ungrammatical	0.333	0.225	0.375	0.267	0.653	0.694
Total	0.389	0.164	0.497	0.211	0.694	0.196
Control						
Grammatical	0.307	0.207	0.417	0.317	0.400	0.211
Ungrammatical	0.200	0.172	0.217	0.209	0.267	0.196
Total	0.253	0.147	0.317	0.188	0.333	0.152

guistic group differed significantly from the recast and control groups. There was also a tendency toward a significant difference between the recast and control groups.

The analysis of the grammatical and ungrammatical items showed no significant group differences on the immediate posttest for either the grammatical,  $F(2, 34) = 0.853$ ,  $p = .436$ , or ungrammatical items,  $F(2, 34) = 0.753$ ,  $p = .480$ . However, on the delayed posttest, there were significant differences on both the grammatical,  $F(2, 34) = 6.697$ ,  $p < .01$ , and ungrammatical items,  $F(2, 34) = 4.769$ ,  $p < .05$ , with the metalinguistic group differing significantly from the control on both. Additionally, the metalinguistic group differed significantly from the recast group on the grammatical items, with a similar trend toward significance for metalinguistic over recasts on the ungrammatical items.

### Untimed Grammaticality Judgment Test

The descriptive statistics for regular past tense on the grammaticality judgment test (see Table 4) show relatively high levels of accuracy on the pretest, ranging from 69% to 78%. These accuracy scores generally increased over both posttests.

The ANCOVAs show overall that there is no difference for group on the immediate posttest,  $F(2, 34) = 0.714$ ,  $p = .498$ , although there is for the delayed posttest,  $F(2, 34) = 4.493$ ,  $p < .05$ . The post hoc contrasts for the delayed posttest showed that the metalinguistic group differed significantly from the recast group and that there was a trend toward significance for metalinguistic over the control.

**Table 4.** Grammaticality judgment test results

Groups	Pretest		Immediate posttest		Delayed posttest	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Recast						
Grammatical	0.714	0.122	0.833	0.147	0.784	0.142
Ungrammatical	0.854	0.129	0.844	0.152	0.813	0.146
Total	0.784	0.062	0.839	0.071	0.798	0.075
Metalinguistic						
Grammatical	0.738	0.134	0.929	0.114	0.941	0.074
Ungrammatical	0.844	0.108	0.833	0.154	0.844	0.094
Total	0.791	0.073	0.881	0.115	0.892	0.065
Control						
Grammatical	0.586	0.247	0.786	0.181	0.871	0.142
Ungrammatical	0.788	0.145	0.813	0.189	0.738	0.190
Total	0.687	0.135	0.799	0.098	0.805	0.096

The ANCOVAs do not reveal any group differences on the immediate posttest for the grammatical  $F(2, 34) = 1.482, p = .243$ , or ungrammatical items  $F(2, 34) = 0.092, p = .912$ . Additionally, there were no differences on the delayed posttest for the ungrammatical items,  $F(2, 34) = 0.900, p = .417$ . However, there were significant differences on the delayed posttest for the grammatical items,  $F(2, 34) = 5.194, p < .05$ , with the post hoc contrasts showing that the metalinguistic group differed significantly from the recast group, and the control differed significantly from the recasts.

### The Metalinguistic Test

The results from the metalinguistic test (see Table 5) show that all three groups had high accuracy scores on the pretest and that these generally remained

**Table 5.** Metalinguistic test results

Groups	Pretest		Immediate posttest		Delayed posttest	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Recast	0.958	0.144	0.833	0.326	1.00	0.000
Metalinguistic	0.833	0.246	0.917	0.194	0.917	0.194
Control	0.850	0.241	0.900	0.210	0.850	0.337

**Table 6.** Imitation: new/old item results

Groups	Pretest		Immediate posttest		Delayed posttest	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Recast						
New	0.226	0.247	0.286	0.220	0.345	0.232
Old	0.250	0.309	0.467	0.287	0.583	0.301
Metalinguistic						
New	0.321	0.203	0.373	0.230	0.631	0.223
Old	0.483	0.232	0.657	0.265	0.783	0.233
Control						
New	0.271	0.171	0.271	0.228	0.271	0.196
Old	0.210	0.242	0.380	0.239	0.420	0.148

high over the two posttests. Due to the small number of test items ( $n = 2$ ), inferential statistics were not calculated for the metalinguistic test.

### Old versus New Items

Tables 6 and 7 present the descriptive statistics for test performance on past tense verbs that appeared in the treatment tasks (i.e., old items) and for past tense verbs that were not the object of feedback (i.e., new items). If the effect of the treatment were only evident on the old items, this would suggest that corrective feedback caters only to item learning; if the effect could be shown

**Table 7.** Grammaticality judgment test: new/old item results

Groups	Pretest		Immediate posttest		Delayed posttest	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Recast						
New	0.823	0.113	0.854	0.139	0.802	0.125
Old	0.750	0.184	0.821	0.138	0.792	0.116
Metalinguistic						
New	0.740	0.135	0.875	0.131	0.906	0.094
Old	0.857	0.122	0.881	0.147	0.869	0.129
Control						
New	0.613	0.092	0.800	0.121	0.750	0.156
Old	0.786	0.226	0.800	0.154	0.857	0.117

to extend to new items, this would constitute evidence of generalization (i.e., system learning).

The results of the ANCOVAs reveal that there were significant differences between the groups on the pretest oral imitation scores. Once these differences are taken into account, there were no differences for the new items on the immediate posttest,  $F(2, 34) = 0.397, p = .676$ ; however, there were differences on the delayed posttest,  $F(2, 34) = 8.943, p < .01$ . The post hoc contrasts show that the metalinguistic group is significantly more accurate than both the recast and control groups. Similarly, for the old items, there was no difference on the immediate posttest,  $F(2, 34) = 1.211, p = .312$ , but there was on the delayed posttest,  $F(2, 34) = 3.188, p = .056$ , with the post hoc contrasts showing that the metalinguistic group was significantly more accurate than the control group. Thus, the results reported for the oral imitation test as a whole apply equally to old and new items.

The results of the ANCOVAs for the grammaticality judgment test reveal that there were significant differences between the groups on the pretest scores. Once these differences are taken into account, there was no difference among the groups for the old items on either the immediate posttest,  $F(2, 34) = 0.452, p = .640$ , or the delayed posttest,  $F(2, 34) = 0.817, p = .451$ , nor for the new items on the immediate posttest,  $F(2, 34) = 0.467, p = .632$ . However, there was a significant difference for the new items on the delayed posttest,  $F(2, 34) = 4.295, p < .05$ , which showed the metalinguistic group to be significantly more accurate than the recast group as well as a trend toward metalinguistic group over the control group.

## Summary

Table 8 summarizes the main results, focusing on the statistically significant differences in the pairwise comparisons.

## DISCUSSION

An inspection of the pretest oral imitation test scores suggests that all of the learners initially had only limited implicit knowledge of past tense *-ed*. This was especially apparent in their inability to produce the correct forms when asked to imitate and correct sentences containing errors in structure. In contrast, the grammaticality judgment pretest scores were high (i.e., above 70% in the two experimental groups). It is also noticeable that scores on the ungrammatical sentences were higher than on the grammatical sentences. If the ungrammatical sentences are taken as affording a better measure of explicit knowledge than the grammatical sentences, as suggested by R. Ellis (2004) and demonstrated in R. Ellis (2005), this might explain the higher scores on the ungrammatical sentences. Also, the metalinguistic test indicated a high

**Table 8.** Summary of statistically significant differences

Test	Pretest	Immediate posttest	Delayed posttest
Oral imitation			
Overall	Significant group differences	No significant differences	M > C (old and new items) M > R (new items) R > C (trend)
Grammatical	Significant group differences	No significant differences	M > C, R
Ungrammatical	Significant group differences	No significant differences	M > C M > R (trend)
Grammaticality judgment			
Overall	Significant group differences	No significant differences	M > C (trend) M > R (new items)
Grammatical	Significant group differences	No significant differences	M, C > R
Ungrammatical	Significant group differences	No significant differences	No significant differences

*Note.* M = metalinguistic group; R = recast group; C = control group.

level of explicit knowledge of past tense *-ed*. Thus, the pretest scores can be interpreted as showing that the learners generally possessed a high level of explicit knowledge of past tense *-ed* but were lacking in implicit knowledge.

The descriptive statistics in Table 3 and the results of the ANCOVAs show that the corrective feedback resulted in significant differences among the groups on the oral imitation test for past tense *-ed*, but these differences were only evident on the delayed posttest. Corrective feedback also led to gains on the grammaticality judgment test. However, the gains were almost entirely due to improved performance on the grammatical sentences, which R. Ellis (2004, 2005) argued tap more into implicit knowledge.

These results suggest that corrective feedback has an effect on the learning of implicit knowledge. Indeed, overall, the feedback appears to have had a greater effect on the learners' implicit knowledge than on their explicit knowledge, although this might simply reflect the fact that the learners possessed ceiling levels of explicit knowledge at the beginning of the study. It is possible, of course, that the treatments increased learners' awareness of the grammatical targets of the oral imitation test, thus encouraging them to monitor their output using their explicit knowledge. However, we do not believe that this occurred. First, when asked at the end of the final test if they were aware of which grammatical structures the test was measuring, only one learner was able to identify past tense. Second, as Table 9 shows, there is no clear evidence that the experimental groups were monitoring more than the control group or more in the posttests than in the pretest. If the learners were attempting to use their explicit knowledge in this test, we would have expected a much higher incidence of self-correction.<sup>3</sup>

Further evidence that the corrective feedback induced changes in learners' implicit knowledge can be found in the fact that the effects of the experimental treatments on the oral imitation test scores were more evident 2 weeks after the instruction than 1 day after. This finding reflects previous research (e.g., Mackey, 1999), which has also shown that the effects of instruction become more apparent in delayed tests that tap the kind of language use likely to measure implicit knowledge. The enhanced accuracy evident in the oral imitation delayed posttest is indicative of the learners' successful incorporation of the target structure into their interlanguage systems.

**Table 9.** Number of instances of participant self-correction during the oral imitation test

Group	Pretest	Immediate posttest	Delayed posttest
Metalinguistic	3	1	4
Recasts	1	7	2
Control	2	1	3

The main purpose of the study was to investigate the relative effects of explicit and implicit corrective feedback on the acquisition of both types of knowledge. In this study, explicit corrective feedback was operationalized as metalinguistic information, and implicit corrective feedback was operationalized as recasts. The results point to a distinct advantage for metalinguistic information despite the fact that the learners in the recast group received substantially more corrective feedback than those in the metalinguistic group (see Table 2). Nor was the advantage found for the metalinguistic group only evident in the grammaticality judgment test: It was also clearly evident in the oral imitation test. Also, metalinguistic feedback (but not recasts) was found to result in learning that generalized to verbs not included in the treatment, which suggests that system learning took place.

How can we explain the general superiority of explicit feedback over implicit feedback? In the earlier discussion of theoretical issues relating to corrective feedback, we noted that in connectionist models of L2 acquisition, explicit corrective feedback in the context of communicative activity can facilitate the conversion of explicit knowledge into implicit knowledge.<sup>4</sup> Explicit feedback is more likely than implicit feedback to be perceived as overtly corrective, as the examples in (6) and (7) suggest. In both episodes, the teacher's feedback move overlaps with the learner's preceding move, but because the metalinguistic feedback is longer (six words as opposed to one), it might have been better attended to and perceived as overtly corrective. In both episodes, however, the learner successfully repairs the error following the feedback move, but, again, there is evidence of greater awareness that repair is needed in the metalinguistic episode. Whereas in (6) the learner simply repeats the reformulated past tense verb, in (7) the learner's "yes" seems to overtly acknowledge that repair was required. Thus, metalinguistic feedback—in comparison to recasts—seems more likely to lead to greater depth of awareness of the gap between what was said and the target norm, thereby facilitating the acquisition of implicit knowledge. It is also important to recognize that the metalinguistic feedback, as illustrated in (7), does not intrude unduly in the communicative flow of the activity. It constitutes a brief time-out from communicating, which allows the learner to focus explicitly but briefly on form. The effectiveness of the metalinguistic feedback, therefore, might derive in part from the high level of awareness it generates and in part from the fact that it is embedded in a communicative context.

(6) L: yesterday two boys, Joe and Bill visit their rich uncle=  
 T: = Visited  
 L: Visited their rich uncle.

(7) L: yesterday Joe and Bill ah went to ah Bill's grandmother and visit their grandmother=  
 T: = and visit > you need past tense  
 L: Visited, yes.  
 Key: = signifies overlapping elements; > signifies rising intonation; L = Learner;  
 T = Teacher

The superiority of the metalinguistic feedback only reached statistical significance in the delayed oral imitation and grammaticality judgment posttests. However, gains from pretest to the immediate posttest were also evident. Thus, the general pattern of the results was: pretest scores < immediate posttest scores < delayed posttest scores; that is, the benefits of the metalinguistic feedback became more evident as time passed. This finding supports the claims of R. Ellis (1994) and N. C. Ellis (2005) that explicit L2 knowledge can enhance the processes involved in the development of implicit knowledge (e.g., noticing and cognitive comparison); that is, the awareness generated by metalinguistic feedback promotes the kind of synergy between explicit and implicit knowledge that is hypothesized to underlie L2 learning.

The relatively weak effect found for either type of feedback on the ungrammatical sentences in the grammaticality judgment test reflects the fact that the learners possessed the explicit knowledge required for judging such sentences from the beginning, which was clearly evident from their high pretest scores on the ungrammatical sentences on the grammaticality judgment test and near-perfect scores on the metalinguistic test.

One final comment is in order. All of the learners in this study demonstrated partial implicit knowledge of past-tense *-ed*, as demonstrated by their performance on the oral imitation pretest. It is possible that for corrective feedback of any kind to have an effect on learning, the structures must be at least partially established in the learners' interlanguages. Further research is needed to establish whether corrective feedback is effective in enabling learners to acquire completely new grammatical structures.

## CONCLUSION

This study demonstrates that explicit feedback in the form of metalinguistic information is, overall, more effective than implicit feedback (in the form of recasts) and contributes to system as well as item learning. Table 10 summarizes the actions that learners are hypothesized to carry out in order to process feedback for acquisition (based on Carroll's, 2001, account of corrective feedback), and the extent to which the two types of feedback engage these processes. It illustrates how both implicit and explicit types of feedback might facilitate these actions and it demonstrates why explicit feedback might do so more effectively than implicit feedback. In particular, explicit feedback seems more likely to promote the cognitive comparison that aids learning.

The study reported in this paper incorporated a number of unique methodological features. To the best of our knowledge, this is the first experimental study that has compared the effects of online explicit corrective feedback in the form of metalinguistic explanations and implicit corrective feedback in the form of recasts in classroom-based instruction. Other studies that have compared explicit and implicit corrective feedback (see Table 1) have either examined other forms of corrective feedback or have been laboratory based,

**Table 10.** Facilitative potential of implicit and explicit feedback compared

General action	Specific step	Implicit feedback (recasts)	Explicit feedback (metalinguistic information)
Detecting the error	The learner's attention must switch from meaning to form; this can only occur if the corrective intention is recognized, which happens only if the feedback is perceived as irrelevant to the ongoing discourse.	Partial recasts are more likely to induce a switch in attention from meaning to form than full recasts; however, the corrective force of partial recasts is not self-evident, as repetitions serve other functions in discourse.	Metalinguistic information is clearly irrelevant to the ongoing discourse, which makes a switch from meaning to form likely and the corrective intention of the feedback clear.
	The learner must be able to locate the error.	Recasts make the location of the error clear.	Repeating the error and then giving a metalinguistic clue makes the location of the error clear.
Correcting the error	The learner must be able to distinguish the phonological dissimilarity between his or her erroneous form and the target form in the feedback.	Recasts make it possible for the learner to compare two phonological forms (the erroneous form and the target form), but the learner has to attend to the difference.	The process of comparing the erroneous and target phonological form is enhanced because the teacher repeated the incorrect form before supplying metalinguistic information.
	The learner must be able to decide whether pronunciation, morphology, syntax, or semantics is causing the problem. The learner can make use of explicit knowledge to decide.	Recasts provide no clues as to the nature of the error; that is, the learner is left to infer whether the error is one of pronunciation or morphology.	The nature of the error is made clear. A connection is made between the implicit use of a specific form (the error) and an explicit representation of the target form required.

*Note.* Recasts refer to partial recasts of the kind used in this study.

or both. We argue that metalinguistic explanation and recasts constitute the best exemplars of explicit and implicit corrective feedback, as both are supported by previous research that shows them to be effective in promoting learning. We also argue that from a pedagogical perspective, it is important to examine corrective feedback within the classroom context. We do not believe that it is easy to extrapolate the results obtained from laboratory studies that involve one-on-one interactions to classrooms in which the teacher interacts with the whole class. In our view, ecological validity can only be achieved through classroom-based research.

Another unique feature is that the corrective feedback occurred in the context of learners performing communicative tasks. In Long's (1991) terms, therefore, the feedback constituted focus-on-form. In the majority of studies listed in Table 1, the feedback was part of a focus-on-forms. The importance of the study reported in the preceding sections is that it demonstrates that when learners receive explicit feedback on their attempts to communicate, acquisition takes place. Thus, the study supports Long's claims about focus-on-form. In the view of the teacher who taught the lessons and of the researchers who inspected the transcripts of the corrective feedback episodes, the learners were engaged with performing the tasks and the primary focus was on meaning, not form. The study demonstrates that metalinguistic feedback does not detract unduly from the communicative flow of a lesson.

A third unique feature of the study is to be found in the instruments used to measure learning. These measures were based on prior research designed to develop relatively separate measures of implicit and explicit knowledge (R. Ellis, 2005; Erlam, *in press*). These measures enable us to consider whether the effects of the corrective feedback were related to implicit as well as explicit knowledge. The results reported earlier suggest that corrective feedback results in gains in implicit knowledge. As such, they constitute evidence against the claims that have been advanced by some theorists (Krashen, 1982; Schwartz, 1993) that corrective feedback plays no part in the development of implicit knowledge.

As in all classroom studies, there are inevitable limitations. First, the sample size for this study was small. Also, we were forced to use intact groups, with the result that the groups were not equivalent at the commencement of the study, thus obligating the use of ANCOVA. Second, because our main aim was to compare the relative effectiveness of the two types of corrective feedback, we only included a testing group as a control group (i.e., we did not have a control group that completed the communicative tasks without any corrective feedback). Third, the length of the treatments was very short (approximately 1 hr). It is possible that with a longer treatment, recasts would have proved more effective. Fourth, the structure we chose for study was a structure that the learners had already begun to acquire. In one respect, this can be considered a strength, as it enabled us to examine which type of corrective feedback works best for structures already partially acquired. However, in another respect, it constitutes a weakness, in that we are unable to say whether corrective feedback (and what type of corrective feedback) is effective in establishing new knowledge. As always, further research is needed.

## NOTES

1. Other studies have examined the relationship between implicit/explicit feedback and learner uptake (e.g., Oliver & Mackey, 2003), but these are not included in this review, which focuses exclusively on the effects of feedback on L2 acquisition as measured in posttests. The extent to which uptake constitutes a measure of acquisition is controversial, with many researchers, including ourselves, preferring to view it as evidence of noticing.

2. Rosa and Leow's (2004) study also involved other experimental conditions. These depended on whether there was a prior grammatical explanation of the target feature. The feedback conditions summarized in Table 1 did not include any prior explanation.

3. When scoring the oral imitation test, self-corrections were discounted.

4. A reviewer of a draft version of this paper pointed out that the results could be explained in terms of the learners' having automatized their declarative knowledge of past tense *-ed* as a result of the treatment. This interpretation draws on the distinction between declarative and procedural knowledge, which informs skill-building theories of the kind advocated by DeKeyser (1998). However, we have chosen to frame the paper in terms of the implicit/explicit distinction, noting with Eysenck (2001) that recent changes in the definitions of both pairs of terms have brought them closer together, making it "increasingly difficult to decide on the extent to which different theories actually make significantly different predictions" (p. 213).

## REFERENCES

- Ayoun, D. (2004). The effectiveness of written recasts in the second language acquisition of aspectual distinctions in French: A follow-up study. *Modern Language Journal*, 88, 31–55.
- Carroll, S. (2001). *Input and evidence: The raw material of second language acquisition*. Amsterdam: Benjamins.
- Carroll, S., Roberge, Y., & Swain, M. (1992). The role of feedback in adult second language acquisition: Error correction and morphological generalizations. *Applied Psycholinguistics*, 13, 173–198.
- 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, 357–366.
- DeKeyser, R. M. (1993). The effect of error correction on L2 grammar knowledge and oral proficiency. *Modern Language Journal*, 77, 501–514.
- DeKeyser, R. M. (1998). Beyond focus on form: Cognitive perspectives on learning and practicing second language grammar. In C. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 42–63). New York: Cambridge University Press.
- DeKeyser, R. M. (2003). Implicit and explicit learning. In C. Doughty & M. Long (Eds.), *The handbook of second language acquisition* (pp. 313–348). Oxford: Blackwell.
- Doughty, C. (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., & 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.
- Dulay, H., & Burt, M. (1974). A new perspective on the creative construction process in child second language acquisition. *Language Learning*, 24, 253–278.
- Ellis, N. C. (1994). Introduction: Implicit and explicit language learning—an overview. In N. C. Ellis (Ed.), *Implicit and explicit learning of languages* (pp. 1–31). San Diego, CA: Academic Press.
- Ellis, N. C. (2005). At the interface: How explicit knowledge affects implicit language learning. *Studies in Second Language Acquisition*, 27, 305–352.
- Ellis, R. (1994). A theory of instructed second language acquisition. In N. C. Ellis (Ed.), *Implicit and explicit learning of languages* (pp. 79–114). San Diego, CA: Academic Press.
- Ellis, R. (2003). *Task-based language learning and teaching*. Oxford: Oxford University Press.
- Ellis, R. (2004). The definition and measurement of L2 explicit knowledge. *Language Learning*, 54, 227–275.
- Ellis, R. (2005). Measuring implicit and explicit knowledge of a second language: A psychometric study. *Studies in Second Language Acquisition*, 27, 141–172.
- Ellis, R., & Sheen, Y. (in press). Re-examining the role of recasts in L2 acquisition. *Studies in Second Language Acquisition*.
- Erlam, R. (2005). *Assessing L2 implicit language knowledge*. Manuscript submitted for publication.
- Eysenck, M. (2001). *Principles of cognitive psychology* (2nd ed.). New York: Psychology Press.
- 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. (2003). Factors affecting the success of corrective feedback. In S. Foster-Cohen & A. Nizgorodzew (Eds.), *EUROSLA Yearbook, Volume 1*. Amsterdam: Benjamins.
- Hawkins, R. (2001). *Second language syntax: A generative introduction*. Oxford: Blackwell.
- Iwashita, N. (2003). Negative feedback and positive evidence in task-based interaction: Differential effects on L2 development. *Studies in Second Language Acquisition*, 25, 1–36.

- Kim, H., & Mathes, G. (2001). Explicit vs. implicit corrective feedback. *The Korea TESOL Journal*, 4, 1–15.
- Krashen, S. (1982). *Principles and practice in second language acquisition*. Oxford: Pergamon.
- Leeman, J. (2003). Recasts and second language development: Beyond negative evidence. *Studies in Second Language Acquisition*, 25, 37–63.
- Long, M. (1991). Focus on form: A design feature in language teaching methodology. In K. de Bot, R. B. Ginsberg, & C. Kramsch (Eds.), *Foreign language research in cross-cultural perspective* (pp. 39–52). Amsterdam: Benjamins.
- 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. (in press). Recasts in SLA: The story so far. In M. Long (Ed.), *Problems in SLA*. Mahwah, NJ: Erlbaum.
- Long, M., Inagaki, S., & Ortega, L. (1998). The role of implicit negative feedback in SLA: Models and recasts in Japanese and Spanish. *Modern Language Journal*, 82, 357–371.
- Lyster, R. (1998). Negotiation of form, recasts, and explicit correction in relation to error types and learner repair in immersion classrooms. *Language Learning*, 48, 183–218.
- Lyster, R. (2004). Differential effects of prompts and recasts in form-focused instruction. *Studies in Second Language Acquisition*, 26, 399–432.
- Lyster, R., & Ranta, L. (1997). Corrective feedback and learner uptake: Negotiation of form in communicative classrooms. *Studies in Second Language Acquisition*, 19, 37–66.
- Mackey, A. (1999). Input, interaction and second language development: An empirical study of question formation in ESL. *Studies in Second Language Acquisition*, 21, 557–587.
- Mackey, A., Gass, S., & McDonough, K. (2000). How do learners perceive interactional feedback? *Studies in Second Language Acquisition*, 22, 471–497.
- Mackey, A., Oliver, R., & Leeman, J. (2003). Interactional input and the incorporation of feedback: An exploration of NS-NNS and NNS-NNS adult and child dyads. *Language Learning*, 53, 35–66.
- Makino, T. (1980). Acquisition order of English morphemes by Japanese secondary school students. *Journal of Hokkaido University of Education*, 30, 101–148.
- Muranoi, H. (2000). Focus on form through interaction enhancement: Integrating formal instruction into a communicative task in EFL classrooms. *Language Learning*, 50, 617–673.
- Nagata, N. (1993). Intelligent computer feedback for second language instruction. *Modern Language Journal*, 77, 330–339.
- Nicholas, H., Lightbown, P., & Spada, N. (2001). Recasts as feedback to language learners. *Language Learning*, 51, 719–758.
- Oliver, R., & Mackey, A. (2003). Interactional context and feedback in child ESL classrooms. *Modern Language Journal*, 87, 519–533.
- Panova, I., & Lyster, R. (2002). Patterns of corrective feedback and uptake in an adult ESL classroom. *TESOL Quarterly*, 36, 573–595.
- Philp, J. (2003). Constraints on “noticing the gap”: Nonnative speakers’ noticing of recasts in NS-NNS interaction. *Studies in Second Language Acquisition*, 25, 99–126.
- Rosa, E. M., & Leow, R. P. (2004). Awareness, different learning conditions, and second language development. *Applied Psycholinguistics*, 25, 269–292.
- Sanz, C. (2003). Computer delivered implicit vs. explicit feedback in processing instruction. In B. VanPatten (Ed.), *Processing instruction: Theory, research, and commentary*. Mahwah, NJ: Erlbaum.
- Schmidt, R. (1994). Deconstructing consciousness in search of useful definitions for applied linguistics. *AILA Review*, 11, 11–26.
- Schwartz, B. (1993). On explicit and negative evidence effecting and affecting competence and “linguistic behavior.” *Studies in Second Language Acquisition*, 15, 147–163.
- Sheen, Y. (2004). Corrective feedback and learner uptake in communicative classrooms across instructional settings. *Language Teaching Research*, 8, 263–300.
- Sheen, Y. (in press). Exploring the relationship between characteristics of recasts and learner uptake. *Language Teaching Research*.
- Sorace, A. (1996). The use of acceptability judgments in second language acquisition research. In W. Ritchie & T. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 375–409). San Diego, CA: Academic Press.