

The Effects of the Timing of Corrective Feedback on the Acquisition of a New Linguistic Structure

SHAOFENG LI

University of Auckland
Applied Language Studies and Linguistics
Auckland 1142
New Zealand
Email: s.li@auckland.ac.nz

YAN ZHU

Fudan University
College of Foreign Languages and Literatures
Shanghai 200433
People's Republic of China
Email: yan_zhu@fudan.edu.cn

ROD ELLIS [affiliated with two institutions]

University of Auckland
Applied Language Studies and Linguistics
Auckland 1142
New Zealand

Shanghai International Studies University
Shanghai 200083
People's Republic of China

Email: r.ellis@auckland.ac.nz

ABSTRACT

The article reports a study investigating the comparative effects of immediate and delayed corrective feedback in learning the English past passive construction, a linguistic structure that the learners had little prior knowledge of. A hundred and twenty EFL learners from four intact classes at a Chinese middle school were randomly assigned to four conditions: immediate feedback, delayed feedback, task-only, and control. The three experimental groups attended a two-hour treatment session where they performed two dictogloss (narrative) tasks in groups, each followed by a reporting phase in which they took turns to tell the narrative to class. The two feedback groups received either immediate or delayed corrective feedback in the form of a prompt, followed by recasts of utterances containing errors in their use of the target structure. No effect for the corrective feedback was found on elicited imitation test scores but both the immediate and delayed feedback resulted in gains in grammaticality judgement test scores, with immediate feedback showing some advantage over delayed feedback. We interpret these results as showing that the feedback only aided the development of declarative/explicit knowledge and that the advantage found for immediate feedback was due to the learners using the feedback progressively in the production of new past passive sentences whereas this did not occur in the delayed feedback condition.

Keywords: feedback timing; immediate feedback; delayed feedback

Research on oral corrective feedback (CF) spans a period of almost forty years. Early research (e.g., Allwright, 1975; Chaudron, 1977) was descriptive in nature, documenting the various strategies that teachers used to correct errors, and while there has been continued interest in the discourse features of CF (e.g., Seedhouse, 2004), much of the more recent research has been experimental in nature, driven in part by a practical need to identify the most effective way of correcting L2 learners' errors and by theoretical issues concerning the role of negative evidence, input and output in L2 acquisition.

Four recent meta-analyses that have synthesized experimental CF studies (Li, 2010; Lyster & Saito, 2010; Mackey & Goo, 2007; Russell & Spada, 2006) all reported sizable effect sizes, indicating that CF assists acquisition. The meta-analyses also investigated a number of variables that moderate the effect that CF has on acquisition – the research setting (laboratory versus classroom), the target feature (pronunciation, vocabulary, grammar, pragmatics), the type of feedback (implicit versus explicit; input-providing versus output-prompting), the length of the treatment (short, medium, long), the durability of the effect (immediate versus delayed post-test scores), learner variables (e.g., age) and the types of outcome measures (comprehension versus production; constrained versus free response). Notably missing from this list of moderating variables, however, is the timing of the feedback – that is, whether correction took place immediately (i.e., 'online' more or less as soon as an error occurred) or was delayed (i.e., took place after the oral activity that served as the context for the correction was completed). The reason for this omission is that, to date, there have been insufficient studies investigating the timing of CF to make meta-analysis possible.

The purpose of the study reported in this article is to contribute to the ongoing research on oral CF by investigating the effect of timing. We compare the effect of correcting past passive errors while learners were performing communicative tasks with the effect of correcting errors committed in the performance of the same communicative task but after the

task was completed. Below, we consider both the practical and theoretical grounds for investigating the timing of CF, the limited second language research that has investigated the timing variable, our choice of corrective feedback strategy (corrective recasts), and the types of tests we used to measure learning.

PRACTICAL AND THEORETICAL ISSUES

In a seminal article, Hendrickson (1978) addressed five central questions about corrective feedback in language pedagogy, one of which was when learner errors should be corrected. In the Audiolingual Method, teachers were expected to correct immediately after an error had been committed in order to prevent bad habits becoming entrenched. However, this practice was challenged by advocates of the humanistic methods that gained popularity in the 1970s. These emphasized positive or non-judgemental assessment in order to “promote a positive self-image of the learner as a person and language learner” (Ur, 1996, p. 243). The advent of communicative language teaching in the 1980s led to a distinction being drawn between ‘fluency’ and ‘accuracy’ (Brumfit, 1984) and the instructional practices involved in each. Teachers were often advised to avoid immediate correction during ‘fluency’ work. Hedge (2000), for example, observed that teachers’ notes accompanying course books frequently instruct teachers to leave correction until the end of fluency activities. Scrivener (2005) suggested that teachers should make a list of the errors that occurred during a fluency activity and address them when the activity was over. Immediate correction during accuracy work, however, was not just approved of but strongly recommended. Thus, the general position in mainstream language pedagogy was “if the objective is accuracy, then immediate correction is likely to be useful; if the aim is fluency, then lengthy, immediate correction that diverts from the flow of speaking is less appropriate” (Scrivener, 2005, p. 299).

Theoretical positions in second language acquisition (SLA), however, view immediate corrective feedback during fluency work as not only desirable but as especially

facilitative of interlanguage development. One claim, derived from Long's (1996) Interaction Hypothesis and the importance he attached to Focus on Form (Long, 1991), is that immediate feedback consisting of recasts (i.e., reformulations of erroneous learner utterances) creates a "window of opportunity" (Doughty, 2001, p. 257) for learners to carry out a cognitive comparison between their own attempt to express a particular meaning and the correct target language form required to do so. This enables learners to construct a form-function mapping that can potentially be incorporated into their interlanguage systems. From this perspective, recasts are seen as especially beneficial because they are non-obtrusive, they provide learners with the correct form in a brief time-out from semantic processing, and they facilitate the noticing-the-gap that Schmidt (1994) argued is necessary for acquisition to take place. Immediate correction by means of recasts, so the argument goes, provides learners with the kind of input data needed for the development of linguistic competence (i.e., implicit L2 knowledge).

Immediate correction also receives support from a different theoretical perspective. According to Transfer-Appropriate Processing (TAP), "we can use what we have learned if the cognitive processes that are active during learning are similar to those that are active during retrieval" (Lightbown, 2008, p. 27). Lightbown went on to suggest that this can explain why linguistic forms learned in isolation are not available for use in communicative interaction. When learners experience using linguistic forms communicatively, deeper processing occurs than when attention is purely on surface form. In immediate CF, learners are required to process receptively and/ or productively those linguistic forms needed to express what they want to say, which activates learning processes in a context where they are primarily focused on communicating. In accordance with TAP, the learning that takes place in such a context will be available for subsequent use in spontaneous communicative events. In contrast, in delayed CF there is a temporal disjunction between the act of communicating

and the receiving of correction so that the processing involved does not occur under real operating conditions, thus encouraging metalinguistic understanding rather than the development of true linguistic competence.

TAP provides a theoretical rationale for immediate correction consisting of both output-prompting (e.g., requests for clarification or elicitation) and input-providing types of CF such as recasts. That is, it does not make any claims about which type of CF is most beneficial. Lyster (2004) drew on skill-learning theory, which proposes that acquisition begins with a declarative representation of a linguistic feature, which is at first reinforced through mechanical type practice exercises and then proceduralized and eventually automatized through production practice under real operating conditions (DeKeyser, 2007). Lyster argued the case for prompting learners to self-correct while they are communicating on the grounds that this would facilitate the proceduralization of those grammatical forms that learners had already partially acquired. His study provided evidence that prompts were more effective than recasts. However, it is difficult to see how prompts can assist learners' acquisition of new grammatical features. In this case, input-providing input (as in recasts) is surely needed.

All three theoretical frameworks support the immediate provision of CF of one type or another but they do not explicitly address delayed CF. However, given the emphasis they place on the contextualized processing of CF and given that delayed CF is necessarily decontextualized (i.e., it occurs outside the communicative events in which the errors were committed), these theories can be construed as disfavouring delayed CF.

Quinn (2014), however, refers to theories in cognitive psychology that provide a rationale for delayed CF. Preparatory Attention and Memory Theory proposes that learning is enhanced when complete attention is given to the task at hand rather than divided between multiple tasks (McDaniel et al., 1998). It also suggests that when learners perform a difficult

task they will realize they lack crucial information and thus are more likely to pay attention to this information when it is subsequently made available to them. In this way the link between retrospective and prospective memory is established making learning possible. This theory suggests that learners might be better off just focusing on meaning when they perform a challenging task rather than dividing their attention between meaning and form as happens when they process CF, and thus correcting errors is better delayed until they have completed the task. Reactivation and Reconsolidation Theory (Nader, 2003) also suggests that delayed feedback may be preferable. Research based on this theory shows that when memories are reactivated in conditions that make them susceptible to change, their labile state allows for reconsolidation. Reconsolidation is retrieval-induced and occurs in both declarative and procedural memory systems. Quinn notes that both immediate and delayed CF can instigate the retrieval and reconsolidation of linguistic forms but argues that delayed CF is better equipped to do so because it allows more time for both retrieval and reconsolidation to take place.

To sum up, several SLA theories point to an advantage for immediate CF, namely that it ensures the activation of those cognitive mechanisms responsible for the development of linguistic competence (i.e., implicit knowledge). Delayed feedback may only aid only the development of explicit knowledge. Thus, the SLA theories dispute the advice of language educators that feedback should be delayed in 'fluency' work. In contrast, theories from cognitive psychology do offer a rationale for such advice. They suggest that delayed CF may be preferable because it removes the need for concurrent attention to both meaning and form and because it is better able to foster the processes of retrieval and reconsolidation involved in the development of both declarative and procedural memories.

COMPARATIVE STUDIES OF IMMEDIATE AND DELAYED CORRECTIVE FEEDBACK

Hattie and Timperley (2011) in their conceptual analysis of feedback studies in education distinguished feedback about task, where the feedback indicates whether performance on a task is correct or incorrect, and feedback about process, where the feedback addresses the students' choice of strategies for accomplishing a task. Reviewing the substantial body of educational research on feedback, they concluded that feedback about task is more effective when it is immediate, especially when the learning task involved is an easy one. In contrast, delayed feedback is more beneficial for addressing process issues. Arguably, feedback on language entails feedback about task rather than process. It is directed at concrete items (i.e., the specific errors that learners make) rather than how a task is approached strategically. For these reasons, the educational research points to an advantage for immediate feedback in language learning.

In contrast to the interest in the timing of feedback in educational research, there has been very little attention paid to it in SLA. The focus has been almost exclusively on immediate feedback with very few studies of delayed CF and even fewer that have compared the effects of immediate and delayed CF. An interesting descriptive study by Rolin-Ianzati (2010) investigated how two teachers of L2 French went about providing delayed feedback following a role-play activity. She identified two different approaches that correspond to the input-providing and output-prompting types of feedback found in immediate CF. That is, in one approach the teacher provided the corrections while in the other the teacher elicited corrections from the students. The teachers were systematic in the approach they chose, with one teacher preferring the first approach, and the other teacher the second for most corrections. Rolin-Ianzati's did not investigate whether the type of delayed feedback had any effect on learning. However, it was useful as a guide to how input-providing delayed feedback could be given.

Written CF is more or less invariably delayed as it is provided after learners have

completed a piece of writing. Research on delayed written CF shows that it is effective in helping learners achieve greater accuracy in new writing (Bitchener & Ferris, 2012) although there is uncertainty as to whether it contributes to the development of linguistic competence (Williams, 2012). Two written CF studies compared the effects of immediate and delayed feedback. Lavolette et al (2015) reported no difference in accuracy in new writing between a group of L2 learners who received immediate and delayed feedback. However, in this study the ‘immediate’ feedback was only provided after students had completed writing, not while they were writing. Aubrey & Shintani (2014), however, did compare feedback given while the students were writing – delivered by means of Google Docs – and delayed feedback given shortly after they had finished writing and reported that in a delayed posttest the immediate feedback resulted in greater accuracy in the use of the grammatical structure targeted in this study (hypothetical conditionals).

Aljaafreh & Lantolf’s (1994) often-cited study investigated the oral CF directed at errors that individual learners had made in a piece of writing. In accordance with sociocultural theory, the researchers argued that to be effective the feedback needs to be graduated (i.e., fine-tuned to the learner’s developmental level by identifying the least explicit type of correction needed to enable a learner to self-correct an error). They provided evidence to suggest that over time less explicit feedback was needed to enable students to self-correct their errors, which they saw as evidence of learning. The learners first completed a piece of writing and then were asked to read aloud their written text and were stopped and corrected whenever an error occurred. In effect, then, the study points to the effectiveness of immediate oral feedback.

Two studies (Siyari, 2005; Varnosfadrani, 2006) compared the effect of immediate and delayed oral CF, reporting no difference in their effect on learning. However, both studies suffered from major design problems making it difficult to reach any conclusion with

confidence. In Varnosfadrani's study, for example, CF type (implicit versus explicit) was confounded with CF timing (immediate versus delayed). A third study (Quinn, 2014) was well-designed. In this laboratory-based study 90 intermediate-level adult ESL learners were randomly assigned to immediate, delayed, and no CF conditions. The grammatical target was English passive constructions. In a pretest – immediate posttest – delayed posttest design involving an aural grammaticality judgement test, an oral written error correction test and a written error-correction test, the learners completed three communicative tasks. The immediate and delayed feedback consisted of a prompt that pushed the learners to self-correct followed by a recast. There were statistically significant improvements resulting from all three conditions but there were no statistically-significant differences between conditions. In other words, not only was the timing of the feedback of no significance but, in this study, feedback itself conferred no learning advantage.

Thus, whereas the educational research reviewed in Hattie & Timperley (2007) demonstrated an advantage for immediate feedback directed at task goals, the SLA research, meagre as it is, has failed to find any difference in the effect of immediate and delayed feedback on L2 acquisition. The education research examined feedback in relation to 'knowledge' and 'understanding' in content areas of the school curriculum whereas the SLA research investigated feedback on learners' use of language while they were primarily focused on communicating. It is possible that feedback—including the timing of the feedback—functions differently depending on whether learners are engaged in intentional knowledge-learning (as in the education research) or incidental skill-learning (as in the SLA studies).

CHOICE OF CORRECTIVE FEEDBACK STRATEGY

Much of the SLA research on corrective feedback has been directed at investigating the relative effects of different CF strategies. Ellis & Shintani (2014) classified these strategies in terms of whether they are input-providing (e.g. recasts and explicit correction) or

output-prompting (e.g., clarification requests and elicitation) and also according to whether the corrective force of the feedback is explicit (e.g., explicit correction and elicitation) or implicit (e.g., recasts and clarification requests). Research investigating these different strategies has been motivated by both practical and theoretical concerns and has been surrounded by considerable controversy as reflected in Goo & Mackey's (2013) defence of recasts and Lyster & Ranta's (2013) response. The results of research to date are not conclusive. Overall explicit CF appears to be more effective than implicit CF, but there is plenty of evidence to show that both recasts and prompts are effective. Summing up the research, Lyster, Saito, and Sato (2013) argued "a variety of CF types is probably more effective than consistent use of only one type" and "it may not be necessary or even possible for researchers to identify the single most effective CF strategy" (p. 21).

A potentially ideal CF strategy—the type of feedback provided in the instructional treatment of this study—is a hybrid corrective move consisting of a prompt that performs the dual function of drawing the learner's attention to the problematic nature of the production and eliciting self-correction, and a recast that provides the correct form in the absence of self-correction. This hybrid feedback package, which is called corrective recast by Doughty and Varela (1998), encourages the learner to retrieve the correct form from his/her linguistic repertoire and scaffolds the learner's performance when the need for assistance arises.

MEASURING LEARNING

Of central importance to any study investigating the effects of corrective feedback or any other type of form-focused instruction is the distinction between explicit and implicit knowledge. Ellis (2005) defined explicit knowledge as knowledge that is conscious, metalinguistic and typically only available for use in controlled processing. Implicit knowledge is knowledge that is tacit, proceduralized, and thus available for automatic processing. Ellis (2005) reported a psychometric study investigating whether tests could be

designed to provide relatively separate measures of these two types of knowledge. The results of the study suggested that an oral elicited information test could be used to measure L2 learners' implicit knowledge and a grammaticality judgment test (in particular, the ungrammatical sentences in this test) afforded a measure of explicit knowledge. A criticism of form-focused instruction research is that all too often learning has been measured by means of discrete point tests or grammaticality judgement tests (Norris & Ortega, 2000), which bias learners to the use of their explicit knowledge of grammatical rules. Drawing on these results, we included an oral elicited imitation test and a grammaticality judgement test in the study reported below in order to investigate whether immediate and delayed CF had differential effects on the two types of knowledge.

RESEARCH QUESTIONS

The study aimed to contribute to research on corrective feedback in two main ways: by comparing the effect of immediate and delayed CF and by investigating the effect of CF on the acquisition of a grammatical structure that the learners had no prior knowledge of. The study sought to answer the following research questions.

RQ1. Does immediate and delayed CF result in the acquisition of explicit of the English past passive construction?

RQ2. Does immediate and delayed CF result in the acquisition of implicit knowledge of the English past passive construction?

RQ3. Is there any difference in the effect of immediate and delayed CF on the acquisition of (a) explicit and (b) implicit knowledge of the English past passive construction?

METHOD

Participants

The participants were 120 eighth-grade EFL learners at a public school in South-Eastern China. They were aged between 13 and 15, with an average age of 14.1 years. The

learners, who were later divided into four groups, were from four intact classes, with class sizes ranging from 55 to 60. These four classes were randomly selected from a total of 18 classes at the eighth grade. Each of the resultant groups consisted of 30 learners and, depending on the type of treatment they received, the groups were labelled immediate feedback, delayed feedback, task-only, and control, respectively (details for each treatment condition are provided subsequently). One-way ANOVAs detected no significant differences between the four groups in their mid-term exam scores, $F(3, 111) = .09, p = .97$, pretest scores on the grammaticality judgment test, $F(3, 110) = .47, p = .70$, or pretest scores on the elicited imitation test, $F(3, 98.59) = 1.60, p = .20$.

The learners attended six 40-minute English lessons on a weekly basis. A longitudinal observational study carried out by one of the authors (Zhu, 2015) at this school revealed that the learners primarily received traditional grammar-based and teacher-fronted instruction, with limited exposure to communicative teaching. All these participants had Chinese as their native language. The learners reported that they had studied English for 4 to 11 years, with an average length of 6.1 years.

The instructor for the treatment sessions was one of the researchers who, at the time of this study, was a Ph.D. student with 11 years of EFL teaching experience. She had worked at this research site for more than three years on another project and thus was familiar with the instructional setting, but she had not taught these learners prior to this study.

Target Structure

The English past passive was chosen as the target structure for two reasons. First, the students had not received any prior instruction on this structure. According to the curriculum of the school, which was based on the national curriculum mandated by the Chinese Ministry of Education, the passive construction is introduced in the second semester of the eighth grade, several months after the current study. Second, the passive voice is a late-

acquired and linguistically complex feature (see Quinn, 2014). It is argued (Ju, 2000; Qin, 2008) that the English passive voice is especially difficult for Chinese EFL learners because of the lack of inflectional morphology in Chinese and the difference in the formation of the passive voice between the two languages. Whereas the canonical structure of the English passive voice is “Subject (receiver of action) + Be + Past Participle + By + Agent (performer of action)”, in Chinese it is “Subject (receiver of action) + Bèi (passive marker, similar to ‘by’) + Agent (performer of action) + Verb”. For example, if translated into English, the Chinese version of the passive sentence “He was arrested by the police” would be “He by police arrest *le* (aspect marker)”. An interview with four eighth-grade English teachers prior to the study also indicated that the passive voice was one of the most challenging structures for students at this level. The focus of the study was only the form of the past passive, and no attempt was made to teach other forms of the passive voice.

Instructional Treatment

Tasks. The learners in the three experimental groups received two hours of instruction consisting of two consecutive sessions with a five-minute break. In each of the two sessions, the learners completed a dictogloss task in which they listened to a narrative presented by PowerPoint and read by the instructor, worked in pairs practicing retelling the narrative to each other, and were then called on individually to tell the story to the rest of the class. The decision to use tasks of the same type was motivated by research that points to the benefits of task repetition in alleviating learners’ processing burden and facilitating attention to linguistic forms (e.g., Gass, et al, 1999).

The steps for implementing the tasks were as follows:

- a) Preparation. At the beginning of each treatment session, the instructor went through a list of words on a PowerPoint slide that would appear in the subsequent narrative task together with the students. Then the instructor asked two brainstorming questions

relating to the topic of the task to arouse the learners' interest and activate their schematic knowledge.

- b) Presentation of input materials. After the preparation, the learners listened to a narrative told by the instructor three times. First, the instructor told the narrative orally to provide the learners with some initial understanding of the content. Then the instructor showed the narrative on PowerPoint, each slide containing one or two sentences with vocabulary annotations. The instructor read the slides aloud and, after finishing each slide, paused for five seconds before moving on to the next slide so that the learners had an opportunity to process the input material. The learners were provided with a list of cue words—nouns and verbs from the narrative (e.g., child—injure) – that could be consulted during the presentations of the narrative and which the learners could refer to later in the pair work. Finally, the instructor repeated the narrative a third time orally to consolidate the learners' knowledge of the content and familiarize them with the language involved.
- c) Pair work. After listening to the narrative three times, the students worked in pairs to practice retelling the narrative to each other, and they were also asked to add an ending to the narrative which was to be shared later with the rest of the class.
- d) Public reporting. After the pair work, the students were called on to tell the narrative to the class in collaboration with their partners, with one student in each pair telling part of it before handing it over to his/her partner. The partner then completed the story and reported the ending they had agreed on earlier.
- e) Wrap-up. The task cycle ended with a brief whole-class discussion of which group's story ending was the most interesting.

This study involved two feedback groups that performed two dictogloss tasks following the steps previously described and received corrective feedback on their non-targetlike use of

the English past passive, a comparison group (task-only) that did not receive feedback and only performed the tasks, and a control group that did not receive any treatment and only took the pretests and posttests. The two feedback groups differed in terms of when they received feedback: The immediate feedback group received online feedback as they performed the narratives whereas the delayed feedback group received offline feedback after the tasks were completed. Including a task-only group makes it possible to distinguish the effects of feedback from the effects of performing communicative tasks alone.

Input Materials. Two narrative texts (see Appendix A) were used in the treatment, one for each dictogloss task. One, which was composed by the researchers, was a report of a car accident, and the other was a story about an earthquake in Haiti, based on a story from *Reader's Digest*. The two texts were seeded with 30 cases of past passive (15 each), out of which 18 involved regular verbs and 12 irregular verbs. We consulted the textbooks for the learners and two experienced local teachers when developing the materials. These texts related to topics the learners were familiar with and were comparable in length with those in the textbook. All the vocabulary was checked against the current and previous textbooks and the words that the learners had not been exposed to were pre-taught and annotated when the stories were presented on PowerPoint. Prior to the study, the tasks were piloted with a group of eighth-graders ($n = 16$) to fine-tune the linguistic and procedural aspects.

Feedback. Immediate feedback was operationalized as corrective recasts (Doughty & Varela, 1998), where an erroneous utterance was repeated with the error highlighted via prosodic emphasis to encourage self-correction, followed by a recast that reformulated the wrong utterance without altering the meaning. The feedback was implemented as follows:

- a) The teacher repeated the whole sentence, highlighting the passive error with prosodic emphasis and paused 3–5 seconds for self-correction on the student's part. For example,

Student: The driver was arrest.

Teacher: The driver WAS ARREST? (Teacher paused 3–5 seconds for self-correction)

- b) If the student made a self-correction, the teacher confirmed and signalled for the student to continue:

Student: The driver was arrested.

Teacher: Ok. Go ahead.

- c) If the student's response still contained the error or if there was no response at all, the teacher reformulated the whole sentence, highlighting the passive part:

Student: Yes, the driver arrested.

Teacher: The driver WAS ARRESTED (slowly and emphatically).

Delayed feedback was provided after the completion of the second task, and the procedure was the same as that for immediate feedback except that in delayed feedback the teacher initiated the corrective episode by quoting a wrong utterance a learner had produced when performing the tasks. The errors that were corrected were those recorded by the teacher and another researcher who sat in the class during the students' task performance. The delayed feedback was provided as follows:

- a) The teacher quoted a learner's erroneous sentence and asked him/ her to correct it. For example, "Tom, you said 'The driver wanted to run away, but he stopped by a policeman'. Can you say it correctly?" The teacher then paused 3–5 seconds for a response.
- b) If the student self-corrected the teacher stopped there. Otherwise the teacher proceeded to step 3.
- c) The teacher reformulated the sentence, highlighting the passive part, "he WAS STOPPED by a policeman." The teacher then moved on to the next error.

The feedback in both conditions consisted of an output-prompting move that encourages

self-repair followed by an input-providing move that provides positive evidence, thus catering to both Lyster and Ranta's (2013) claim that prompts engage learners in deeper cognitive processing of the linguistic target and to Goo and Mackey's (2013) claim that recasts facilitate the learning of new structures.

While every attempt was made to minimize the differences between the immediate and delayed feedback, the timing of the feedback inevitably led to some differences. For example, it could be argued that prompts differ in the delayed and immediate feedback in that the former constitutes an elicitation and the latter a repetition of the learner's erroneous utterance. However previous research (e.g., Ammar & Spada, 2006; Lyster, 2004; Yang & Lyster, 2010) has treated different prompts as equivalent on the grounds that they all push learners to self-correct. It might also be argued that delayed feedback is more explicit than immediate feedback so that any difference in the effects of the timing is just a reflection of a difference in salience. In fact, efforts were made to ensure that the immediate feedback was salient to the learners through prosodic emphasis. While some recasts are implicit in nature, corrective recasts are clearly more explicit. Also, any difference in the degree of explicitness is an inherent feature of the timing of the feedback. As we pointed out in the introduction, investigating the timing of feedback is of obvious pedagogical significance. We argue that the timing of feedback needs to be investigated in an ecologically valid way and that attempts to rigorously control for inherent differences that arise from the timing of feedback are inappropriate.

Table 1 provides the information about the instructional treatment in the two feedback conditions including the length and instances of feedback treatment, the number of errors, and the number of learners who committed an error. As can be seen, the duration of feedback and number of presenters are very similar between the two groups, although there are slightly more instances of feedback and more errors in the immediate feedback condition.

TABLE 1

Errors and Treatment in the Two Feedback Conditions

Feedback	Duration of CF		Instances of CF		No. of Errors		No. of Presenters	
	Task 1	Task 2	Task 1	Task 2	Task 1	Task 2	Task 1	Task 2
Immediate	497.44	391.86	32	18	38	22	14	9
Delayed	845.8*		19	17	30	20	11	12

Note. *In the delayed feedback condition, CF was provided after both tasks were completed.

Testing

Treatment effects were measured via an untimed grammaticality judgement test and an elicited imitation test, which were intended to provide measures of the learners' explicit and implicit knowledge of the target structure respectively (Li, 2013, 2014). The grammaticality judgment test asked the learners to judge whether an item was grammatical or ungrammatical, and correct the error if it was ungrammatical. The elicited imitation test required each learner to verbally repeat some sentences (grammatical and ungrammatical) presented in an aural mode. Both tests had three versions: pretest, posttest 1, and posttest 2 and the three versions were created by randomly scrambling the same items. Both tests included 30 target items: 20 contained regular verbs and 10 irregular verbs; 20 were old items that targeted verbs from the treatment tasks, and 10 were new items, with all the new items targeting regular verbs. The old items were equally distributed between the two tasks (i.e., 10 each), and the two tasks also contributed an equal number of regular and irregular verbs. To minimize the possibility of the learners transferring their answers from one test to the other, the sentence stimuli in the grammaticality judgment test were different from those in the elicited imitation test, but the contexts for the obligatory use of the target structure were the same in the two tests.

Based on the results of a pilot study where 24 learners from the same cohort

performed the two treatment tasks, three types of errors were included in the ungrammatical items: (1) no “be”, as in “Tony badly injured in a fight with a friend”, (2) bare verb, as in “Two men kill in the accident”, and (3) no past participle, as in “This morning Helen was knock down in the street”. To validate the test items, nine native speakers of English, who were either applied linguists or experienced ESL teachers, were invited to respond to and comment on a large pool of sentences in terms of grammaticality and wording. The piloted items were composed by the researchers using vocabulary from the learners’ textbook and the treatment tasks.

The grammaticality judgement test was developed and presented via an online computer program. The test included 40 items, 30 of which related to the past passive—the target structure, and 10 were distractors relating to structures the learners had been taught, including the third person –s, simple past, comparative adjectives, and prepositions of time. Among the 40 items, 35 were ungrammatical, 5 grammatical, and all the 30 target items were ungrammatical. The use of ungrammatical items in a grammaticality judgement test as a measure of explicit knowledge was based on Gutiérrez’s (2013) finding that learners’ responses to grammatical and ungrammatical items load on separate factors, with the former tapping implicit knowledge and the latter explicit knowledge. In scoring the learners’ answers, one point was given if an ungrammatical sentence was judged to be ungrammatical and the correct form of the passive construction was supplied. The internal reliability for the grammaticality test, indexed by Cronbach’s alpha, was .91 for the pretest and .96 for both posttests.

The elicited imitation test consisted of 40 statements relating to the learners’ personal experience (e.g., “My knee injure on my way to school today”), of which 30 concerned the past passive—the target structure, and 10 were distractors. Half of the target items ($n = 15$) were grammatical and half ($n = 15$) ungrammatical. Previous research reported no difference

between learners' responses to grammatical and ungrammatical items in an elicited imitation test (Erlam, 2006). The statements were comparable in length (each containing around 10 words) and complexity. The stimuli, read at a normal speed by a native speaker and recorded using a digital recorder, were presented via DMDX, free software used in psycholinguistic studies to present auditory stimuli and to control or record reaction time. During the test, the learners heard the statements one at a time, and after hearing each statement, they were asked to decide whether it was true, not true, or whether they were not sure, and then repeat the sentence in correct English regardless of whether it was true of their personal life/experience. One point was given if a response contained the correct form of the past passive, and because self-corrections may involve the use of explicit knowledge, only first attempts were scored. A reliability analysis showed that the Cronbach's alpha for the elicited imitation test was .68, .76, and .77 for the pretest, the immediate posttest, and the delayed posttest, respectively.

Unlike previous studies (e.g., Ellis et al., 2006) where reaction time was not controlled, thus increasing the chances for learners to access their explicit knowledge, in this study a time limit was imposed on each item. The learners were required to judge the veracity of a sentence and then repeat it within the time allocated after which the program moved on to the next item regardless of whether the previous item was responded to. Because the items in the test varied in length, the reaction time allocated for each item was based on the average of the time taken for each item by 26 learners from classes other than those selected for the main study took to complete that item.

Procedure

The study spanned three days. On Day 1, the learners took the grammaticality judgment and elicited imitation pretest. One week later, on Day 2, the experimental groups received the instructional treatment, followed by the immediate posttests while the control

group just completed the tests. On Day 3, two weeks later, all the learners took the delayed posttests. To minimize the potential modelling effect of a written test on an oral test, the elicited imitation test always preceded the grammaticality judgment test. Throughout the study, all four groups of participants continued with their normal instruction, the learners and their teachers were never informed of the purpose of the project, and the teachers did not provide any instruction on the target structure.

Analysis

We elected to both investigate total accuracy scores for past passive and accuracy scores separately for regular and irregular past participles in past passive constructions. The decision to examine regular and irregular past particles separately was motivated by research that has shown that the acquisition of these differs. Ullman (2001), for example, proposed that irregular past verb forms are lexical and thus stored in declarative memory (i.e., they are explicit) whereas regular forms can eventually be computed in procedural memory (i.e., they involve the implicit system). From this perspective, the clearest test of whether the feedback contributed to the learners' implicit knowledge was the effect it had on passive constructions involving the regular past participle.

To prevent misleading conclusions resulting from the influence of extreme values, outliers in each group in each test were identified and removed prior to analysis. To identify outliers, raw scores were transformed into z scores (standardized scores for which the mean is zero and standard deviation is 1), and any score 2.5 standard deviation units above or below the mean was considered an outlier. To explore whether the learners' test scores varied as a function of the type of treatment they received and the timing of tests, a mixed design repeated measure analysis of variance (ANOVA) was conducted, with group (treatment type) as the between-group variable, and scores on the pretest and the two posttests the within-group variable. If the initial mixed design ANOVA detected significant main or interaction

effects, one-way ANOVAs were conducted to locate the source of significance, followed by post hoc pairwise comparisons. Because the assumption of homogeneity of variances of ANOVA was violated in most cases, the *F* test was performed using the Brown–Forsythe adjustment, and the post hoc comparisons were conducted with the Dunnett T3 corrections. In pairwise comparisons, in addition to using the *p* value to evaluate the significance of a difference in mean scores, the effect size in the form of Cohen’s *d* was calculated to show the magnitude of the difference. *d* is primarily based on mean difference and thus overcomes the limitation of the sole reliance on the results of null-hypothesis significance testing (i.e. the *p* value), which is sensitive to sample size and prone to Type II error. Following Cohen’s benchmarks for interpreting effect sizes, .2, .5, and .8 were considered as small, medium, and large effects respectively.

RESULTS

Grammaticality Judgment Test Results

Overall. The descriptive statistics appear in Table 2 while the group means are graphically displayed in Figure 1. These results show that (1) the learners’ pretest scores were very low (the group means ranged 1.29–1.89 out of 30), suggesting that they possessed almost no knowledge of the English past passive, (2) there was very little between-group variation in the learners’ pretest scores, (3) all learners performed better on the posttests than the pretest, and (4) the experimental groups showed higher scores than the control group on both posttests.

TABLE 2

Means and Standard Deviations for Grammaticality Judgment Test

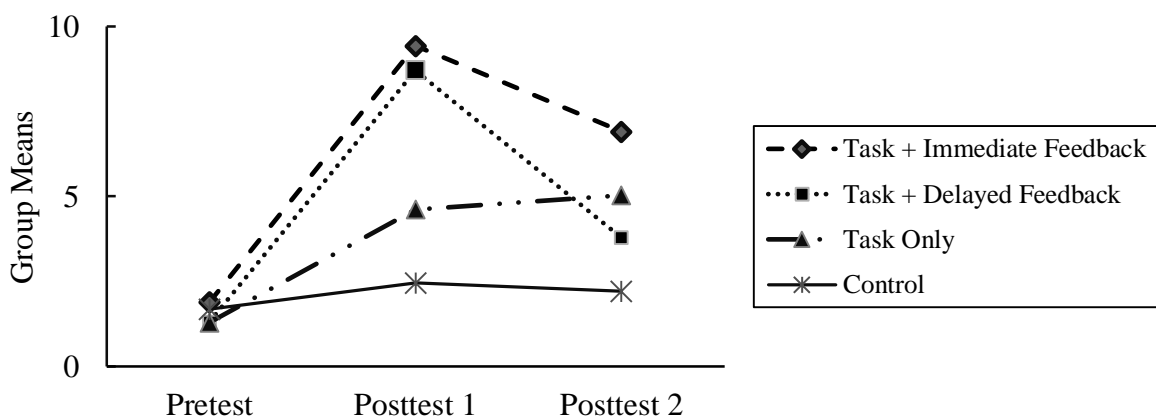
Group	Pretest			Posttest 1			Posttest 2		
	<i>n</i> ^a	<i>M</i> ^b	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>
Immediate Feedback	28	1.89	3.30	30	9.43	8.19	29	6.90	8.20

Delayed Feedback	29	1.31	1.79	30	8.70	9.45	28	3.79	4.10
Task-Only	28	1.29	1.80	29	4.62	5.69	29	5.03	6.79
Control	29	1.69	1.98	28	2.46	2.63	29	2.21	2.48

Note. a. The number of participants ranges from 28 to 30 across groups and tests after outliers were removed; b. the highest possible (full) score is 30.

FIGURE 1

Group Means on Grammaticality Judgment Test over Time



The repeated measure ANOVA showed a significant effect for time, $F(2, 214) = 40.84, p < .00$, for group, $F(3, 107) = 3.13, p = .03$, and for time \times group interaction, $F(6, 214) = 57.32, p < .00$. These results indicated that the scores of the learners were different across the three time points, that overall there were differences between the four groups' test scores regardless of time, and that the performances of the four groups in relation to each other were different at the three time points (for example, the task-only group was outperformed by the delayed feedback group at the time of the immediate posttest but at the time of posttest 2, the reverse was true). Of primary concern is whether significant differences existed between the four groups' posttest scores and whether the differences were due to the difference in the treatment they received. The results of one-way ANOVA revealed that there were no

significant differences among the four groups in their pretest scores (which suggests that any between-group differences in the learners' posttest scores are attributable to the instructional treatments), but there were in their scores on the two posttests. Post hoc pairwise comparisons (see Table 3) found that at the time of posttest 1, both the immediate and delayed feedback groups scored significantly higher than the control group, and the differences were associated with large effect sizes. At the time of posttest 2, immediate feedback was the only treatment group that significantly outperformed the control group. Although the task-only group also showed higher scores than the control group on both posttests, the differences did not reach statistical significance. However, there were no significant differences between the three treatment groups in their scores on either posttest, although the advantage of immediate feedback over task only was near significant ($p = .07$).

TABLE 3
Post Hoc Comparisons for Treatment Effects on Grammaticality Judgment Test*

Group Contrasts	Posttest 1		Posttest 2	
	d^1	p^2	d	p
Immediate Feedback vs. Control	1.11	.00	.74	.03
Delayed Feedback vs. Control	.95	.01	.58	.41
Task-Only vs. Control	.58	.35	.63	.22
Immediate Feedback vs. Delayed Feedback	.02	1.00	.39	.37
Immediate Feedback vs. Task-Only	.59	.07	.17	.92
Delayed Feedback vs. Task-Only	.52	.27	-.23	.95

Note. *The F tests for the posttest scores were performed using the Brown–Forsythe adjustment and the post hoc comparisons were conducted with the Dunnett T3 corrections.
1. effect size (Cohen's d); 2. results of null hypothesis significance testing.

Regular vs. Irregular Verbs. An analysis of the groups' scores for regular and irregular verbs was undertaken to ascertain whether the effects of treatment were different as a function of verb type. The means and standard deviations, organized by group and verb

type, appear in Table 4, which shows that the two feedback groups performed better on regular verbs than irregular verbs on the two posttests.

TABLE 4

Means and Standard Deviations for Grammaticality Judgment Test: Regular vs. Irregular Verbs

Group		Pretest			Posttest 1			Posttest 2		
		<i>n</i> ^a	<i>M</i> ^b	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>
Immediate Feedback	Regular	28	.06	.13	30	.34	.28	29	.34	.28
	Irregular	29	.07	.10	28	.21	.21	28	.15	.21
Delayed Feedback	Regular	29	.04	.08	30	.31	.33	29	.31	.33
	Irregular	29	.04	.05	30	.25	.28	28	.08	.09
Task-Only	Regular	28	.04	.07	28	.14	.17	28	.14	.17
	Irregular	29	.07	.10	29	.14	.19	29	.13	.21
Control	Regular	29	.06	.08	28	.09	.11	29	.09	.11
	Irregular	29	.04	.06	29	.06	.09	29	.05	.07

Note. a. The number of participants ranges from 28 to 30 across groups and tests after outliers were removed; b. percentage scores were used because of the unequal numbers of regular ($n = 20$) and irregular ($n = 10$) verbs in the test.

Mixed design ANOVAs showed that, for both regular and irregular verbs, there were significant effects for time and group and for time \times group interaction (see Appendix B for details). For both verb types, one-way ANOVAs did not detect significant differences between the four groups in their pretest scores, but there were significant between-group differences in their posttest scores. The learners' posttest scores were subjected to post hoc pairwise comparisons to investigate the comparative effects of the different treatment types on regular and irregular verbs. The results reported in Table 5 show that at the time of posttest 1, both feedback groups performed significantly better than the control group for both regular and irregular verbs, and the mean score for immediate feedback was

significantly higher than that for task-only in the learning of regular verbs. At the time of posttest 2, the only significant difference was found between immediate feedback and control in the learning of regular verbs. On both posttests, the two feedback groups (in comparison with control) showed larger effect sizes for their performance on regular verbs than irregular verbs.

TABLE 5

Post Hoc Comparisons for Regular and Irregular Verbs on Grammaticality Judgment Test*

Group Contrasts	Posttest 1				Posttest 2			
	Regular		Irregular		Regular		Irregular	
	d^1	p^2	d	p	d	p	d	p
Immediate Feedback vs. Control	1.13	.00	.75	.01	.76	.03	.48	.11
Delayed Feedback vs. Control	.93	.01	.88	.01	.63	.25	.38	.64
Task-Only vs. Control	.45	.83	.57	.23	.59	.43	.58	.27
Immediate Feedback vs. Delayed Feedback	.03	.99	-.27	.99	.25	.76	.27	.48
Immediate Feedback vs. Task-Only	.77	.01	.31	.72	.27	.65	.06	1.0
Delayed Feedback vs. Task-Only	.63	.08	.54	.42	.03	1.0	-.19	.77

Note. *The F tests for the posttest scores were performed using the Brown–Forsythe adjustment and the post hoc comparisons were conducted with the Dunnett T3 corrections.
1. effect size (Cohen’s d); 2. results of null hypothesis significance testing.

Old vs. New Items. An analysis was conducted to investigate whether the treatment effects included new items or were restricted to old items. Recall that the grammaticality judgment test included 30 items, out of which 20 concerned verbs that appeared in the treatment tasks and 10 involved new verbs. Among the 20 old items, 10 targeted regular verbs and 10 irregular verbs; the 10 new verbs were all regular. Given the difference in the learners’ performance for regular and irregular verbs, the analysis was only conducted for regular verbs. The descriptive statistics reported in Table 6 showed that overall there was little difference in the learners’ performance on the old and new items.

TABLE 6

Means and Standard Deviations for Grammaticality Judgment Test: Old vs. New Regular Verb Items*

Group		Pretest			Posttest 1			Posttest 2		
		<i>n</i> ^c	<i>M</i> ^d	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>
Immediate Feedback	Old ^a	28	.46	.96	30	3.40	2.81	28	2.18	2.45
	New ^b	28	.79	1.68	30	3.40	3.01	29	2.62	3.13
Delayed Feedback	Old	29	.31	.66	30	2.90	3.02	28	1.21	1.47
	New	28	.39	.78	28	3.30	3.74	29	1.96	2.32
Task-Only	Old	28	.50	1.00	29	1.32	1.70	29	1.93	2.43
	New	28	.25	.58	29	1.62	2.18	29	1.72	2.47
Control	Old	29	.45	.69	29	.97	1.40	29	.62	1.05
	New	29	.72	1.19	28	1.07	1.27	29	1.03	1.29

Note. a. Items that appeared in the treatment tasks; b. new items; c. the number of participants ranges from 28 to 30 across groups and tests after outliers were removed; d. the highest possible (full) score is 10.

The one-way ANOVA did not detect significant differences between the four groups' pretest scores for either old or new items. Mixed design ANOVAs found that for both old and new items, there were significant main effects for time and group and significant interactions between time and group (Appendix B). One-way ANOVAs and post hoc T-tests (Table 7) showed that at the time of posttest 1, both immediate and delayed feedback outperformed control on both old and new items while immediate feedback worked better than task only on old items. At the time of posttest 2, only immediate feedback performed significantly better than control in the learning of old items, and there were no other significant differences.

TABLE 7

Post Hoc Comparisons for Old and New Items on Grammaticality Judgment Test*

	Posttest 1		Posttest 2	
	Old	New	Old	New

Group Contrasts	d^1	p^2	d	p	d	p	d	p
Immediate Feedback vs. Control	1.08	.00	.97	.00	.82	.02	.64	.09
Delayed Feedback vs. Control	.88	.02	.90	.02	.57	.41	.62	.33
Task-Only vs. Control	.20	.95	.57	.81	.67	.06	.59	.71
Immediate Feedback vs. Delayed Feedback	.12	.98	-.09	1.0	.40	.39	.13	.39
Immediate Feedback vs. Task-Only	.90	.01	.47	.07	.12	.99	.13	.99
Delayed Feedback vs. Task-Only	.71	.09	.50	.21	.02	.69	.00	.69

Note. *The F tests for the posttest scores were performed using the Brown-Forsythe adjustment and the post hoc comparisons were conducted with the Dunnett T3 corrections.

1. Effect size (Cohen's d); 2. results of null hypothesis significance testing.

Elicited Imitation Test Results

The elicited imitation test aimed to gauge the learners' implicit knowledge of the target structure. The means and standard deviations for each group over time are displayed in Table 8, and the means are plotted on the graph in Figure 2. It can be seen that the learners' scores were low on the pretest as well as the two posttests, with the group means ranging from 1.19 to 4.10, and that all groups improved from the pretest to the posttests

A repeated measure ANOVA revealed a significant effect for time, $F(2, 200) = 41.42$, $p < .00$, but there was no significant effect for group (treatment), $F(3, 100) = .92$, $p = .81$, or for Time \times Group interaction, $F(6, 200) = .92$, $p = .43$. One-way ANOVAs found no difference in the groups' pretest scores, $F(3, 109) = 1.57$, $p = .20$, in their scores on the immediate posttest, $F(3, 114) = .25$, $p = .86$, or in their scores on the delayed posttest, $F(3, 116) = 1.20$, $p = .31$. The significant time effect was due to the increase in the learners' overall performance from the pretest to the posttests. These results suggest that the improvement in the learners' posttest scores in comparison with their pretest scores was attributable to practice effects and that there were no effects for the instructional treatments. The data were subjected to further analysis to explore the impact of other variables such as

regular vs. irregular and old vs. new, but the analyses failed to detect any significant treatment effects.

TABLE 8

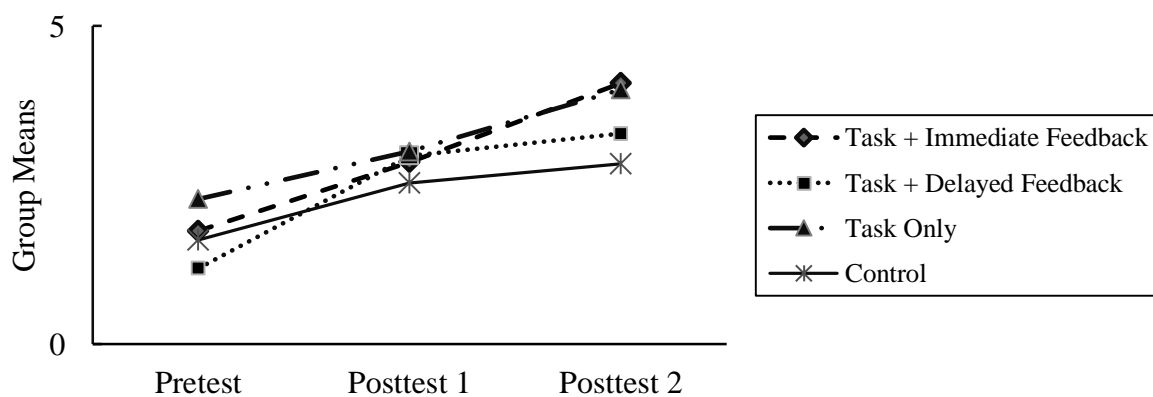
Means and standard deviations for elicited imitation test

Group	Pretest			Posttest 1			Posttest 2		
	<i>n</i> ^a	<i>M</i> ^b	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>
Immediate Feedback	27	1.78	1.87	28	2.86	2.26	29	4.10	3.45
Delayed Feedback	26	1.19	1.49	28	2.96	2.67	29	3.31	3.12
Task-Only	29	2.28	2.28	29	3.03	2.43	30	4.00	2.85
Control	28	1.64	1.68	30	2.53	2.36	29	2.83	2.33

a. The number of participants ranges from 28 to 30 across groups and tests after outliers were removed; b. the highest possible (full) score is 30.

FIGURE 2

Group Means on Elicited Imitation Test over Time



DISCUSSION

We begin by considering what the results show about the effect of the feedback on the learners' implicit and explicit knowledge of the target structure (past passive). We then

provide answers to the research questions by examining the differential effects of the immediate and delayed corrective feedback on the learners' explicit knowledge.

Effects of the Instruction on Learners' Explicit and Implicit Knowledge

The two tests used in the study—the grammaticality judgement test (GJT) and the elicited imitation test (EIT)—were intended to provide measures of learners' explicit and implicit knowledge of the target structure (past passive). The results of these two tests are clear-cut. The immediate and delayed corrective feedback resulted in statistically significant gains in the GJT posttests but not in the EIT posttests. If these tests are taken as measures of explicit and implicit knowledge respectively, the conclusion is that the feedback contributed to the learners' explicit knowledge but not their implicit knowledge. This result differs from some other studies that have investigated the effects of recasts (e.g., Mackey & Philp, 1998; Révész, 2012), which report results showing that recasts lead to improved accuracy in the kinds of language use (e.g., free oral production) likely to tap implicit knowledge. Why, then, did the corrective recasts in this study not lead to improved performance in the EIT?

There are a number of possible answers to this question. One is that the EIT failed to capture gains in implicit knowledge. The test differed from the EIT used in other studies (e.g., Zhang, 2015) in that it was administered via computer rather than face-to-face and that the response time for each sentence was restricted. The aim was to make it difficult for learners to draw on their explicit knowledge, but it is possible that the EIT was simply too demanding on these learners' processing abilities to capture any changes that had taken place in their implicit knowledge. A second explanation is that these learners were not developmentally ready to acquire the target structure. The past passive is morphologically complex (Williams & Evans, 1998) and difficult to acquire. The third explanation is that the instruction provided in this study was insufficient to effect development of implicit knowledge. It consisted of two tasks completed in the same lesson lasting only two hours. Arguably, more intensive

instruction is needed to have any effect on learners' implicit knowledge of a difficult grammatical structure such as past passive. The final explanation is that the learners in this study were accustomed to learning explicitly, so that even though no explicit instruction was provided, they used the input from the tasks and the feedback provided to construct a metalinguistic representation of the past passive, and that the effort they put into achieving this blocked the incidental and implicit processes that lead to implicit knowledge. This last explanation can also account for why the instruction had a very clear effect on the learners' acquisition of explicit knowledge, as measured by the GJT. We will now turn to consider the research questions but only in terms of the effect that the two kinds of feedback had on the learners' explicit knowledge.

Is Immediate Feedback Effective?

The results showed that on posttest 1, the learners receiving immediate feedback outperformed the control group in the GJT, regardless of their previous knowledge, of whether the verbs were regular or irregular, and of whether the test items were old or new. However, the effects of the immediate feedback were less evident in posttest 2.

Much of the previous research involving immediate feedback (e.g., Ellis et al, 2006; Yang & Lyster, 2010) has investigated the effect of recasts on structures that learners already possessed some prior knowledge of and reported results that show it was effective. Doughty & Varela (1998) reported that corrective recasts—the type of recast used in this study—were also effective in improving learners' accurate use of past tense. However, Long (2015) argued that recasts are needed for the acquisition of new linguistic features and went on to list their advantages: They provide feedback that (1) is contextualized and motivating because it is the learner's message or linguistic performance that is at stake, (2) is contingent and caters to the learner's internal syllabus, (3) displays the error and the correct form in juxtaposition so the learner immediately notices the gap, and (4) alleviates the learner's processing burden

and increases the chances for a focus on form by only modifying the incorrect part while maintaining the meaning, and (5) “capitalizes on a symbiotic relationship between explicit and implicit learning, instruction, and knowledge” (p. 317). Goo and Mackey (2013) have also emphasized the need to “examine L2 targets to which learners have never been exposed” (p. 153). This study constitutes a start in this direction.

The pretest showed that the learners had no or only very limited knowledge of the past passive. Immediate feedback in the form of corrective recasts helped them to learn this structure and to some extent maintain this learning over time. However, a caveat is in order. As already noted, the immediate feedback only contributed to the development of the learners’ explicit knowledge. In other words, there was no evidence of the symbiotic relationship between explicit and implicit learning that Long claimed recasts foster.

Is Delayed Feedback Effective?

Learners receiving the delayed feedback scored significantly higher than the learners in the control group on the measure of explicit knowledge when the effects were tested immediately after the treatment, irrespective of learners’ prior knowledge, verb type, and whether the verbs appeared in the treatment. The effects, however, were not sustained after two weeks. These results lend some support to the pedagogic position outlined in the introduction, namely that delaying feedback until learners’ have completed a communicative task is desirable. However, the fact that the effects of the delayed feedback were not sustained suggests that the learning that results from delayed feedback was shallow and that the declarative representations of the past passive that the feedback generated were quickly lost. One anonymous reviewer pointed out that the larger effects of delayed feedback at the time of the immediate posttest may also be due to the closer time proximity between the treatment and the test.

Is There Any Difference in the Effects of Immediate and Delayed Feedback?

This is the core question. Pairwise comparisons of the posttest scores failed to show significant differences between the effects of the immediate and delayed feedback. However, a closer look at the results suggests that the immediate feedback was superior: (1) its effect sizes were larger than those of delayed feedback in 9 out of 10 pairwise contrasts when each feedback type was compared with control and in 8 out of 10 contrasts when the two feedback types were compared with each other, and (2) the significant effects of immediate feedback were sustained after two weeks, while the effects of delayed feedback were only present in the immediate posttest. Again, though, these advantages for immediate feedback were only evident in the GJT; that is they applied only to the learners' explicit knowledge of past passive.

In the introduction we noted that a number of SLA theories lend support to immediate feedback – the Interaction Hypothesis, Focus on Form, Transfer Appropriate Processing and Skill Acquisition Theory. We also noted that theories in cognitive psychology (e.g., Preparatory Attention and Memory Theory and Reactivation and Reconsolidation Theory) predict that delayed feedback will be more effective. By and large the results of this study provide greater support to those SLA theories that support immediate feedback. However, these theories were formulated to explain how learners develop the procedural ability to use those linguistic features they have acquired in free communication (i.e., implicit knowledge). In fact, this study failed to demonstrate that either type of feedback contributed to the development of implicit knowledge. Thus, what is needed is an explanation of why the immediate feedback was more effective for the development of explicit knowledge.

We suggest that the explanation lies in the different processing demands required by the two feedback conditions. It should be noted that there was no difference in the time between the errors and feedback. In both immediate and delayed feedback the feedback was provided immediately following the error. The difference lay in the contextual nature of the

feedback. In the immediate feedback condition, learners received feedback on the errors they had just committed as they struggled to reconstruct the narratives. In the delayed condition, the utterances containing errors were presented one at a time out of context. In the immediate condition, the feedback was linked continuously to the learners' attempts to tell the story and to produce passive sentences. That is, the learners had the opportunity to use the feedback they had received when producing new sentences involving the past passive as they continued to tell the story. In the delayed condition the learners were not required to produce their own sentences, only correct sentences that the teacher presented to them. Skill Acquisition Theory can help to explain why the contextualized condition was more effective. DeKeyser (1998) talked about the importance of learners using their declarative knowledge as a 'crutch' to support their attempts to communicate. In the present study, this did not appear to aid proceduralization of past passive (i.e., no effects on the imitation test), perhaps because the 'practice' afforded by the two tasks was insufficient, but it helped to embed the declarative representations of the target structure more deeply in the learners' memories, which were therefore better sustained over time.

Finally, we note that the results of this study differ from those of other studies that have investigated immediate and delayed feedback—in particular Quinn (2014), which reported no difference in the effect of the timing of feedback on learning. The difference in the results can be traced to the methodological differences between the two studies: (1) Quinn's study was conducted in a laboratory setting, while this study was carried out in the classroom; (2) in Quinn's study a recast was provided regardless of whether the learner was able to self-correct while in this study a recast was only provided when self-repair failed; (3) Quinn provided pre-treatment instruction on the target structure but this study did not; (4) the learners already had considerable prior knowledge of the target structure in Quinn's study but not in this study; (5) in Quinn's study not all feedback in the delayed feedback condition was

delayed because feedback was also provided between tasks, so the so-called delayed feedback for one task served as pre-task instruction for the subsequent task, thus confounding the immediate–delayed distinction. We argue that our study constitutes a methodologically sounder comparison of the effects of the key variable – the timing of the feedback.

We note that the results of our study do mirror those reported in Hattie & Timperley's (2007) synthesis of feedback studies in education research. Hattie and Timperley reported an advantage for immediate feedback directed at informing students whether they were correct or incorrect (what they called 'feedback about task'). We argue that the feedback in this study was directed at factual correctness and that, like the feedback that occurs in subject learning in general, it facilitated the intentional learning of declarative information.

CONCLUSION

This study was undertaken to contribute to the growing body of research on corrective feedback by investigating a variable that has received scant attention to date – the timing of feedback. It also sought to extend current research by investigating the effect of feedback on a grammatical structure that the L2 learners had no or very little prior knowledge of. The study compared the effectiveness of immediate and delayed feedback consisting of corrective recasts on learners' acquisition of English past passive. The pretest showed that the learners had almost no prior knowledge of this structure. Both immediate and delayed feedback proved facilitative of learning this new linguistic structure with some evidence pointing to the superiority of immediate feedback. However, the effects of the feedback were only evident in a GJT indicating that it only contributed to the development of the learners' explicit knowledge. This is, perhaps, not so surprising. The learners had almost no knowledge of the target structure. If the process of learning a structure such as regular past verb forms, which may eventually be processed in procedural memory, involves an initial explicit representation of it, as claimed by Ullman (2001 and N. Ellis (2005) , then, in the short term,

feedback of any kind can only be expected to contribute to such a representation. We found no evidence that the feedback had any effect on the learners' implicit knowledge of the regular past verb forms. The learners would have needed substantially more opportunity to use and receive feedback before they could compute these forms in procedural memory. In other words, our study is perhaps best interpreted as showing that the immediate rather than the delayed feedback assisted the initial stage of acquiring both irregular and regular past participle forms.

The study was motivated by both pedagogical and theoretical concerns. Teachers are often advised to delay providing feedback until learners have completed a communicative task (Scrivener, 2005) on the grounds that providing feedback during the performance of a task will have a negative effect on fluency and on the assumption that delayed feedback is effective. We did not investigate whether the immediate feedback interfered with fluency but just listening to the recordings of the learners' performance of the tasks in the immediate feedback and delayed feedback/ task conditions suggests that it did not. Clearly, though, there is a need to investigate this possibility more thoroughly. The results of the study have shown that delayed feedback is effective – at least where explicit knowledge is concerned – and this lends some support to the claim of teacher trainers such as Scrivener. However, the study also pointed to the superiority of immediate feedback. Perhaps, then, teachers should not feel so constrained by the pedagogic advice they receive regarding the timing of feedback and be prepared to experiment with immediate feedback.

The study does not provide unequivocal support for the SLA theories that point to the need for immediate feedback. These theories seek to account for how feedback contributes to the acquisition of implicit knowledge but this study has only shown that the immediate feedback (and delayed feedback) benefited the development of explicit knowledge. We have suggested that may have been because of the limited nature of the instruction provided or it

might also be because the ‘new’ target structure was too developmentally advanced for the learners. Further research is needed to investigate the theoretical claims advanced by Long (2015) and others regarding the role that recasts play in the acquisition of implicit knowledge of new grammatical features.

ACKNOWLEDGEMENTS

We would like to thank Professor Dingfang Shu from Shanghai International Studies University for his generous support throughout the project. We would also like to extend our sincere gratitude to the students and teachers at Huanan Experimental School in Danyang, Jiangsu Province, China, for their involvement in various aspects of the study. In particular, we would like to acknowledge the following individuals for their support and assistance: Suorong Zhang, Xuping Wang, Wenzhong Wang, Yun Wu, Yufang Yin, Bin Ding, Mingdi Li, Huijun Chen, and Wenjun Zhou.

REFERENCES

- Aljaafreh , A. , & Lantolf , J. (1994). Negative feedback as regulation and second language learning in the zone of proximal development. *Modern Language Journal*, 78, 465–483.
- Allwright, R. (1975). Problems in the study of the language teacher’s treatment of learner error. In M. Burt & H. Dulay (Eds.), *On TESOL ’75: New directions in language learning, teaching, and bilingual education* (pp. 96–109). Washington, DC: TESOL.
- Ammar, A., & Spada, N. (2006). One size fits all? Recasts, prompts, and L2 learning. *Studies in Second Language Acquisition*, 28, 543-574.
- Aubrey, S. & Shintani, N. (2014). The effects of synchronous and asynchronous written corrective feedback on grammatical accuracy in a computer-mediated environment. Paper presented at the meeting of the American Association of Applied Linguistics,

Portland, OR.

Bitchener, J., & Ferris, D. (2012), *Written corrective feedback in second language acquisition and writing*. New York: Routledge.

Brumfit, C. (1984). *Communicative methodology in language teaching*. Cambridge: Cambridge University Press.

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

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. (2007). Skill acquisition theory. In B. VanPatten & J. Williams, *Theories in second language acquisition* (pp. 97-113). Mahwah, New Jersey: Lawrence Erlbaum.

Doughty, C. (2001). The cognitive underpinnings of focus on form. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 206-257). Cambridge: 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.

Ellis, N. (2005). At the interface: Dynamic interactions of explicit and implicit knowledge. *Studies in Second Language Acquisition*, 27, 305-52.

Ellis, R. (2005). Measuring implicit and explicit knowledge of a second language: A psychometric study. *Studies in Second Language Acquisition*, 27, 141-72.

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., & Shintani, N. (2014). *Exploring language pedagogy through second language acquisition research*. London and New York: Routledge.
- Gass, S., A. Mackey, M. Fernandez, & M. Alvarez-Torres. (1999). The effects of task repetition on linguistic output. *Language Learning*, 49, 549-580.
- Goo, J., & Mackey, A. (2013). The case against the case against recasts. *Studies in Second Language Acquisition*, 35, 127-165.
- Gutiérrez, X. (2013). The construct validity of grammaticality judgment tests as measures of implicit and explicit knowledge. *Studies in Second Language Acquisition*, 35, 423-449.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77, 81-112.
- Hedge, T. (2000). *Teaching and learning in the language classroom*. Oxford: Oxford University Press.
- Hendrickson, J. (1978). Error correction in foreign language teaching: Recent theory, research, and practice. *Modern Language Journal*, 62, 387-398.
- Ju, M. K. (2000). Overpassivization errors by second language learners. *Studies in Second Language Acquisition*, 22, 85-111.
- Lavolette, E., Polio, C., & Kahng, J. (2015). The accuracy of competer-assisted feedback and students' responses to it. *Language Learning & Technology*, 19, 50-68.
- Li, S. (2010). The effectiveness of corrective feedback in SLA: A meta-analysis. *Language Learning*, 60, 309-365.
- Li, S. (2013). The differential roles of language analytic ability and working memory in mediating the effects of two types of feedback on the acquisition of an opaque linguistic structure. In C. Sanz & B. Lado, *Individual differences, L2 development & language program administration: From theory to application* (pp. 32-52). Boston, MA: Cengage Learning.

- Li, S. (2014). The interface between feedback type, L2 proficiency, and the nature of the linguistic target. *Language Teaching Research*, 18, 373-396
- Lightbown, P. M. (2008). Transfer appropriate processing as a model for classroom second language acquisition. In Z. H. Han (Ed.), *Understanding second language process* (pp. 27-44). Clevedon, UK: Multilingual Matters.
- Long, M. H. (1991). Focus on form: A design feature in language teaching methodology. In K. de Bot, R. Ginsberg, & C. Kramersch (Eds.), *Foreign language research in cross-cultural perspective* (pp. 39-52). Philadelphia/Amsterdam: John Benjamin.
- Long, M. H. (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: Academic Press.
- Long, M. H. (2015). *Second language acquisition and task-based language teaching*. Malden, MA: Wiley Blackwell.
- Lyster, R. (2004). Different effects of prompts and effects in form-focused instruction. *Studies in Second Language Acquisition*, 26, 399–432.
- Lyster, R., & Ranta, L. (2013). Counterpoint piece: The case for variety in corrective feedback research. *Studies in Second Language Acquisition*, 35, 167-184.
- Lyster, R., & Saito, K. (2010). Oral feedback in classroom SLA: A meta-analysis. *Studies in Second Language Acquisition*, 32, 265-302.
- Lyster, R., Saito, K., & Sato, M. (2013). Oral corrective feedback in second language classrooms. *Language Teaching*, 46, 1-40.
- Mackey, A., & Goo, J. (2007). Interaction research in SLA: A meta-analysis and research synthesis. In A. Mackey (Ed.), *Conversational interaction in SLA: A collection of empirical studies* (pp. 408–452). New York: Oxford University Press.
- Mackey, A., & Philp, J. (1998). Conversational interaction and second language development:

- recasts, responses, and red herrings? *The Modern Language Journal*, 82, 338-356.
- McDaniel, M., Robinson, B., & Einstein, P. (1998). Prospective remembering: Perceptually driven or conceptually driven processes? *Memory & Cognition*, 26, 121-134.
- Nader, K. (2003). Memory traces unbound. *Trends in Neurosciences*, 26, 65-72.
- Norris, J., & Ortega, L. (2000). Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language Learning*, 50, 417-528.
- Qin, J. (2008). The effect of processing instruction and dictogloss tasks on acquisition of the English passive voice. *Language Teaching Research*, 12, 61-82.
- Quinn, P. (2014). *Delayed versus immediate corrective feedback on orally produced passive errors in English*. Unpublished Ph.D. dissertation. University of Toronto.
- Révész, A. (2012). Working memory and the observed effectiveness of recasts on different L2 outcome measures. *Language Learning*, 62, 93-132.
- Rolin-Ianziti, J. (2010). The organization of delayed second language correction. *Language Teaching Research*, 14, 183-206.
- 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. 131-164). Amsterdam: Benjamins.
- Schmidt, R. (1994). Deconstructing consciousness in search of useful definitions for applied linguistics. *AILA Review*, 11, 11-26.
- Scrivener, J. (2005). *Learning teaching: A guidebook for English language teachers* (2nd Edition). Oxford: Macmillan.
- Seedhouse, P. (2004). *The interactional architecture of the language classroom: A conversation analysis perspective*. Malden, MA: Blackwell.
- Siydari, M. (2005). *A comparative study of the effect of implicit and delayed, explicit focus on*

- form on Iranian EFL learners' accuracy of oral production*. Unpublished MA thesis. Iran University of Science and Technology.
- Ullman, M. (2001). The declarative/ procedural model of lexicon and grammar. *Journal of Psycholinguistic Research* 30: 37-69.
- Ur, P. (1996). *A course in language teaching*. Cambridge: Cambridge University Press.
- Varnosfadrani, A. D. 2006. *A comparison of the effects of implicit/ explicit and immediate/delayed corrective feedback on learners' performance in tailor-made tests*. Unpublished PhD dissertation. University of Auckland.
- Williams, J. (2012). The potential role (s) of writing in second language development. *Journal of Second Language Writing*, 21, 321-331.
- Williams, J., & Evans, J. (1998). What kind of focus and on which forms? In C. Doughty & J. Willaims (Eds.), *Focus on form in classroom second language acquisition* (pp. 139-155). Cambridge: Cambridge University Press.
- Yang, Y., & Lyster, R. (2010). Effects of form-focused practice and feedback on Chinese EFL learners' acquisition of regular and irregular past tense forms. *Studies in Second Language Acquisition*, 32, 235-263.
- Zhang, R. (2015). Measuring university-level L2 learners' implicit and explicit knowledge. *Studies in Second Language Acquisition*, 37, 457-86.
- Zhu Y. (2015). *An ethnographic study on foreign language teacher cognition and classroom practices within curriculum innovation in a Chinese secondary school*. Ph.D. dissertation, Shanghai International Studies University.

Appendix A Narrative Texts Used in the Treatment Tasks

1. A Car Accident

There was a bad car accident yesterday. Three people were killed. Also one child was injured. Her leg and arm were broken. Her face was seriously cut. She was driven to the local hospital. Her injuries were treated there. The relatives of the girl were told about the accident.

A witness said, “The car was hit by a big truck. It was badly damaged”. The truck was travelling on the wrong side of the road. The driver of the truck tried to run away. But he was stopped, and he was arrested. He was taken to the police station for questioning. Some bottles of beer were found in his car. He was charged with drunk driving. He was locked in a police cell.

2. An Earthquake

Kiki was raised in a small house in the countryside. One day he was playing when suddenly there was a big earthquake. He was knocked down by the falling bricks. Then the walls fell down. He was trapped in the house. It was very dark. Kiki was badly hurt and could not move. Later Kiki’s mom came back home. She saw the house was destroyed. She thought her boy was buried in the house. She shouted out to him. He could not hear her because he was covered with bricks.

Some dogs were brought to search for him. Kiki was found. The bricks were removed. Kiki was pulled out of the wreckage of the house. He was carried to the local hospital. He was put in an emergency room for treatment. He was given special food to help him recover. He was allowed to leave the hospital after one month.

Appendix B ANOVA Results Different Verb Types and Target Items

	Time		Group		Time \times Group Interaction	
	<i>F</i>	<i>p</i>	<i>F</i>	<i>p</i>	<i>F</i>	<i>p</i>
Regular Verbs	40.63	.00	3.55	.02	5.11	.00
Irregular Verbs	20.86	.00	2.89	.04	3.33	.01
Old Items	39.98	.00	4.30	.01	6.04	.00
New Items	30.24	.00	2.81	.04	2.83	.01