The effects of feedback type and explicit associative memory on the effectiveness of delayed corrective feedback in computer-mediated communication

*Yucel Yilmaz, Indiana University
Gisela Granena, Universitat Oberta de Catalunya
Laia Canals, Universitat Oberta de Catalunya
Aleksandra Malicka, Universitat Oberta de Catalunya

Abstract

The present study examines the impact of the explicitness of corrective feedback and explicit associative memory on the acquisition of -ing/-ed participial adjectives through delayed video-based corrective feedback. Fifty-two L1 Spanish learners were randomly assigned to one of three groups (implicit, explicit, or no-feedback) and performed an interactive task with an experimenter via a video-conferencing tool without receiving any feedback. At the end of the task, the feedback groups received a video replay with inserted oral corrections (either partial recasts or explicit corrections). The no-feedback group performed the interactive task without receiving corrective feedback. A paired-associates test with delayed recall was used to measure explicit associative memory. Pretest-posttest development was measured using oral production and grammaticality judgment tasks. Both corrective feedback groups outperformed the no-feedback group. While no statistical difference emerged between the two delayed corrective feedback groups, a small difference was detected for the explicit group when considering effect sizes. Moreover, a positive relationship was found between explicit associative memory and learning gains on the grammaticality judgment task.

Keywords: Delayed Feedback, Explicit Associative Memory, Corrective Feedback,

Language(s) Learned in This Study: English


Introduction

Broadly speaking, two timing options are available to deliver corrective feedback (CF). Immediate CF is provided shortly after the learner’s non-target-like production, whereas delayed CF is provided after a longer lapse, ranging from the end of the task where the error occurred to several days or weeks after. Despite the overwhelming evidence demonstrating the facilitative role of immediate CF in second language (L2) development (e.g., Li, 2010), language instructors may prefer delayed CF in some language-teaching contexts. It may also be the only option in asynchronous online contexts where instructors and students do not meet synchronously. There is, therefore, a practical need that justifies investigating delayed CF in its own right to determine the conditions under which it can maximize learning opportunities.

* Corresponding Author: Yucel Yilmaz, yyilma@indiana.edu
The present study set out to explore the role of the explicitness of CF (i.e., the extent to which CF includes metalinguistic information) and explicit associative memory (i.e., conscious recollection and storage of events and information) in the effectiveness of delayed CF in a computer-mediated communication (CMC) environment. Beyond its pedagogical value, understanding the role of the explicitness of delayed CF can provide new evidence for the theoretical debate on the relative merits of explicit and implicit CF. Finally, investigating the interaction between explicit associative memory and delayed CF effectiveness is also theoretically important because it can inform about the mechanisms involved in learning from delayed CF.

Corrective feedback in technology-mediated language learning contexts

The cognitive-interactionist perspective, the primary theoretical foundation for CF research, assigns an essential role to attention in second language acquisition (SLA). This view is based on the consensus (Schmidt, 2001; Tomlin & Villa, 1994) that attention to specific linguistic properties (with or without conscious awareness of these properties) is essential for acquisition. Two proposals within the cognitive-interactionist perspective establish a clear theoretical link between attention and CF: the Interaction Hypothesis (Long, 1981) and the focus-on-form approach (Long, 1996). The Interaction Hypothesis (Long, 1981) suggests that meaningful interaction can expose learners to negative evidence through CF and advance L2 learning by directing learners’ attention to linguistic forms. The focus-on-form approach (Long, 1996) claims that success in L2 acquisition is achieved when learners’ attention is momentarily directed to language forms, mainly through CF, during meaning-focused language activities.

Research on oral CF in face-to-face communication has produced abundant evidence demonstrating its positive effect on L2 acquisition (e.g., Li, 2010). CF research in CMC, mostly in the text-based modality, has also shown a positive effect of CF on learning and has attributed its effectiveness to factors such as visual saliency and increased processing time (e.g., Sauro, 2009; Yilmaz, 2012). By contrast, video-based CF has surprisingly received less attention, despite its resemblance to face-to-face interaction, its wide use in distance education, and the fact that it allows keeping the modality of CF and the modality of linguistic performance constant, which could facilitate learning according to transfer-appropriate processing (Lightbown, 2008; Morris et al., 1977).

Only a few studies (e.g., Monteiro, 2014; Saito & Akiyama, 2017; Rassaei, 2017) have focused on the role of video-based CF in SLA. Monteiro (2014) found no difference between metalinguistic feedback, recasts, and a no-feedback control group for English past tense development. Saito and Akiyama (2017) reported gains in English oral language skills such as comprehensibility for both recasts and negotiation, excluding pronunciation accuracy. Rassaei (2017) compared face-to-face and video-based CMC recasts in the acquisition of English articles, with both types outperforming the control group. Overall, these studies suggest variable but potential effectiveness of immediate video-based CF.

Delayed CF and explicit versus implicit feedback

There is emerging evidence in the SLA literature showing that as the time between error and CF increases, the probability of CF resulting in learning gains decreases (Arroyo & Yilmaz, 2018; Li et al., 2016). This inverse relationship has been attributed to the narrow window of opportunity during which CF is effective (Doughty, 2001). However, when it comes to practice, it is not always preferable, or even possible, to implement immediate CF. For example, Harmer (2007) argued that during activities aiming at fluency development, teachers should focus on content and address language issues after the activity is over. Also, in some asynchronous distance language programs, learners can interact synchronously with other learners, but not with instructors (e.g., Canals et al., 2020). Consequently, non-target-like productions cannot be addressed by instructors immediately, necessitating an understanding of when
delayed feedback can be more effective. To address this, a previous study (Canals et al., 2021) examined whether the immediacy of feedback, typically associated with synchronous settings, could be reproduced in asynchronous settings. Specifically, we compared immediate versus delayed CF conditions on the acquisition of -ing/-ed participial adjectives by Spanish L1 learners of English-as-a-foreign language (EFL). Participants completed a communicative task synchronously with an experimenter under one of the two conditions, both of which involved explicit CF. In the immediate condition, CF was provided immediately after a non-target-like form. In the delayed condition, learners were asked to watch a recording of their interaction 24 hours after the task. The recording included feedback instances to previously ignored errors. Both CF groups showed similar learning gains, outperforming the control group, suggesting that delayed CF could be as effective as immediate CF. The study attributed this beneficial effect to the contextualized nature of the delayed CF, which was achieved by having learners watch a video recording of the entire interactive task. A remaining question was whether the effectiveness of delayed CF could vary according to the explicitness or implicitness of the CF.

The distinction between implicit and explicit feedback has been a major focus of research. Although it is difficult to objectively determine the exact level of explicitness of CF, the task is less daunting when the explicitness of a CF type is determined relative to another CF type. Researchers have referred to two types of information to categorize CF on a continuum of explicitness: (a) metalinguistic clues or rules (e.g., “use the past tense”) and (b) information that linguistically highlights any errors made by the learner (e.g., “no”, “that’s wrong”). Accordingly, feedback, including a relatively higher amount of either type of information or both types, has been considered explicit. In contrast, feedback that provides less information in relation to these two categories has been considered implicit (e.g., “No one wants to pay” or “Can you say that again please?” in response to an error like “no one want to pay”). The effectiveness of both implicit and explicit CF has been supported by theoretical claims. Long (2007) argued that implicit feedback in the form of recasts provides information about the L2 precisely when it is needed and thus has a greater likelihood of being perceived as providing negative evidence. However, this type of feedback does not necessarily lead to deeper conscious processing. Because recasts are immediately contingent on learners’ non-targetlike utterances, the juxtaposition of correct and incorrect forms may reduce the involvement of working memory, making it less likely for learners to engage in the type of conscious processing that leads to the understanding of a grammatical rule or regularity, or to awareness at the level of understanding, in Schmidt’s (2001) terms. This quick juxtaposition of the correct and incorrect forms may lead to a beneficial mental process called cognitive comparison (Doughty, 2001). Other researchers, however, have questioned the potential of recasts for L2 learning, arguing that the corrective function of recasts can be ambiguous (Lyster, 1998) or that they do not provide enough metalinguistic information for successful acquisition (e.g., Carroll, 2001). Carroll (2001) claimed that explicit feedback reduces inference and is more likely to succeed than implicit CF.

Despite a substantial body of empirical research comparing implicit and explicit CF, most studies have focused on this issue in the context of immediate feedback provided through face-to-face communication or text-based CMC. Some reported an advantage for explicit feedback (Ellis et al., 2006; Yilmaz, 2012), whereas other studies found no difference between recasts and explicit feedback (e.g., Goo, 2012). In addition, meta-analyses (e.g., Goo et al., 2015; Li, 2010) reported higher effect sizes for explicit feedback. Despite the evidence on the relative effects of immediate explicit versus immediate implicit CF, no studies have examined the relative effects of delayed implicit CF versus delayed explicit CF.

In addition to feedback type, another factor that may interact with the effectiveness of delayed CF is cognitive individual differences (IDs).

Delayed CF and cognitive individual differences

Cognitive IDs can be defined as “aspects of mental functioning, such as memorization and remembering; inhibiting and focusing attention; speed of information processing; and spatial and causal reasoning” (Robinson, 2012, p. 17). They influence the rate and/or the outcomes of L2 learning, either facilitating or
hampering L2 achievement, and their research can inform about the mental processes engaged in L2 learning and about how to enhance L2 learning with targeted pedagogical practices. Many empirical studies have investigated whether and how cognitive IDs are related to the effectiveness of CF (see Granena & Yilmaz, 2018). However, only two studies have investigated the role of cognitive IDs in the effectiveness of delayed CF, both in the context of text-based CMC, and, therefore, their findings have limited value to the present study as the modality investigated differs. Yilmaz and Sağdıç (2019) focused on inhibitory control, whereas Arroyo and Yilmaz (2017) focused on language analytic ability. Neither study showed a statistical interaction between the cognitive ID and the effect of feedback timing on linguistic gains. However, when only the correlations between learning gains and inhibitory control across groups were examined in Yilmaz and Sağdıç (2019), the magnitude of the correlations for the delayed group was the highest. Overall, there is mixed evidence that cognitive IDs can influence the effectiveness of delayed CF. However, given that cognitive IDs have also been found to moderate the effectiveness of CF in immediate conditions, further research on the role of cognitive IDs in delayed CF is clearly warranted.

A cognitive ability that has not been investigated as a moderator of the effectiveness of delayed CF is explicit associative memory. Also known as declarative memory (Cohen & Squire, 1980), explicit associative memory consists of information that is explicitly stored and consciously recalled. Declarative memory involves encoding, storing, and retrieving information with conscious awareness. It can be divided into episodic and semantic memory (Tulving, 1972). Episodic memory is used for learning and retrieving contextualized knowledge of specific events. It contains memories of personally experienced events and the context in which they occurred. Tulving gives examples such as remembering a scheduled appointment, recalling words from a list studied earlier, and autobiographical memories. Semantic memory is memory for general knowledge, concepts, and information that are not tied to a specific learning experience. Learning words initially involves episodic memory, but once learned, the words transition into the semantic store. Paired-associate learning is a memory paradigm used to understand how newly formed associations among stimuli are encoded and retrieved. In a typical paired-associate learning study, participants are asked to learn unrelated word pairs. Later, they are shown one member of each pair as a cue to recall the other.

Language aptitude test batteries such as the MLAT (Carroll & Sapon, 1959) or the LLAMA (Meara, 2005) include a verbal paired associates test and consider explicit associative memory one of the components of aptitude. Testing involves immediate recall measures that test participants immediately after the study phase. Another option, unexplored in the SLA literature, is to test knowledge retention after a time delay. Carroll (1990) already pointed out that tests measuring delayed memory “might be of particular use as measures of language aptitude” (p. 22). Specifically, a delayed recall score showing memory retention abilities could help predict L2 learning outcomes under certain CF conditions. This is because the delayed CF conditions investigated in the present study remind learners of errors made at an earlier stage. In other words, in a delayed CF situation, learners have to re-activate the memory of having made the error since they do not make the error right before receiving feedback. Some learners may be more successful at re-experiencing and retrieving details about past events. If this event involves making an error, a better recall of how the error occurred (i.e., the episode details) could help the learner process the feedback under more advantageous psycholinguistic L2 learning conditions. Therefore, individual differences in explicit associative memory could be one of the factors explaining the extent to which learners benefit from delayed CF.

A relationship between learning outcomes and individual differences in explicit associative memory could inform about the mechanisms involved in learning under the CF conditions presently investigated. If learning from delayed CF is linked to performance on a paired associates learning task, it may be due to a common underlying explicit mental process.
Present Study

The learning benefits of oral CF have been linked to the immediacy that characterizes synchronous communication and to the fact that learners receive CF while vested in the communicative exchange (Long, 1981). The delayed oral CF technique we investigate is similar to immediate oral CF because the CF uses the same modality as the error being corrected, contextually juxtaposing the learner’s non-targetlike utterance and its corresponding targetlike version. Since it is not always feasible to provide CF immediately to L2 learners during oral tasks, research that explores the effectiveness of delayed oral CF is needed.

The literature review above shows that the relative effects of explicit versus implicit CF provided through face-to-face or text-based CMC have been investigated extensively in the context of immediate CF. However, it is not clear whether the results of this research can inform the relative effectiveness of implicit versus explicit CF when CF is provided in a delayed fashion through video-based CMC. Processing conditions in delayed CF could differ fundamentally from those in immediate CF. For example, the absence of interactional pressure may prevent learners’ attention from being divided between comprehending input, contributing to the ongoing interaction, and focusing on CF. Consequently, according to Long (as conveyed to Granena and Yilmaz through personal communications and documented in Granena & Yilmaz, 2022), all undivided attention could be deployed to decode CF and reflect on the target form consciously regardless of the explicitness of the CF.

If learners are predisposed to process CF explicitly and consciously, an indicator of this explicit processing would be the relationship between the effectiveness of CF and the level of a cognitive ability that is commonly associated with explicit processing, such as explicit associative memory. In light of these arguments, this study asked the following research questions:

1. Does the explicitness of delayed feedback impact learners’ linguistic development differently in video-based CMC?

2. Does explicit associative memory differentially affect learners’ linguistic development depending on the explicitness of delayed feedback in video-based CMC?

Method

Design

This study, which was part of a larger-scale research project, followed a pretest-posttest control (no-feedback) group design with random assignment. The final n sizes, after discarding participants who failed to attend one of the scheduled sessions or who did not complete all the tests, were the following: explicit delayed feedback (n = 17), implicit delayed feedback (n = 19), and no-feedback control group (n = 16).

Participants

Fifty-two low-intermediate EFL learners in Spain participated in the study. Their L1 was Spanish. The average age of the participants was 25.7 (SD = 7), and the gender division was 27 females (52%) and 25 males (48%). The study was advertised on several university campuses and online university forums, and the participants were compensated with 50 euros for their participation.

Target structure

English -ing and -ed participial adjectives were the target structures in the study. The use of these participial adjectives has been deemed problematic for learners of English of different levels and L1
backgrounds (Scovel, 1974). These adjectives, which are derived from transitive verbs of state and emotion and formed with the present and past participle of a given verb, are usually presented in grammar books as pairs (e.g., relaxing/relaxed) despite having two different meanings. The adjective “relaxing,” as in “The yoga class was relaxing,” refers to the inanimate noun “the yoga class” being “relaxing.” However, the use of the -ed adjective in the sentence “The girl was relaxed” refers to the feeling of being relaxed as experienced by “the girl”. In other words, in English, the distinction between the two different thematic roles (causer vs. experiencer) is reflected in the adjective. This distinction may not be salient enough for L1 Spanish speakers because, in Spanish, the same adjective form can be used to deliver two different semantic meanings. Consequently, Spanish learners of English have difficulty noticing the semantic distinction of the form-function relationship in the input. Unlike English, in Spanish, the semantic distinction is reflected in the verb (i.e., ser vs. estar). The copula verb estar is used to indicate the experiencer role (e.g., estoy aburrido) and the copula verb ser is used to indicate the causer role (e.g., soy aburrido).

In addition to Canals et al. (2021), which was reviewed above, the only study that investigated the acquisition of English -ing and -ed participial adjectives through CF in the SLA literature was Kim (2002). Participants in Kim (2002) were Korean L1 speakers, and the type of CF used in the study was oral recasts in an immediate condition. Like Spanish, Korean does not distinguish between the two thematic roles (experiencer vs. causer) in the adjective and two different semantic meanings can be delivered with the same form. Kim (2002) found that recasts were effective in facilitating the acquisition of this problematic target structure.

The adjectives that were targeted in the present study were selected from grammar reference books such as Larsen-Freeman and Celce-Murcia (2015). Participial adjectives from emotive verbs that could be easily depicted were first identified and then kept if they were highly frequent in the input and likely to be known by low intermediate learners of English (e.g., bored/boring) and/or had a cognate in Spanish that could avoid lexical confounds (e.g., interested/interesting).

Language measures

A grammaticality judgment task (GJT) and an oral production task (OPT) were used as outcome measures. The OPT was used to measure the ability to retrieve knowledge of the target structure when attention is primarily on meaning, whereas the GJT aimed to measure this ability when attention is primarily on accuracy. Counterbalanced versions (A and B) of each measure were created and administered in three randomized orders to control for potential order effects. Half of the target items in each of the two measures were items from the treatment task. There were no overlaps between the two measures in terms of target items.

In the OPT, participants were asked to compare two situations that were presented on a slide, choosing a form of the word that was displayed at the bottom of the slide (see Appendix A). For example, for a noun such as ‘boredom’, participants produced comparisons such as “I will feel more bored when I listen to a politician”. There were a total of 16 target items and 16 distracters. (The OPT items are available at the IRIS database: https://doi.org/10.48316/zrwgN-2cAKQ.)

In the GJT, participants were given a list of sentences and asked to decide whether each sentence was grammatically correct or incorrect. If the sentence was incorrect, they were asked to underline the error and type in the correction. The GJT included 24 items: eight grammatical, eight ungrammatical, and eight distracters. (The GJT items are available at the IRIS database: https://doi.org/10.48316/K3vZk-tkpKY.)
Cognitive measure

A paired associates test (Karpicke & Roediger, 2008) with delayed recall (a 45-minute interval) was used as a measure of explicit associative memory. The test relies on repeated encoding and retrieval practice to enhance long-term retention. Participants were presented with 40 Swahili-Spanish word pairs visually (e.g., tabibu – doctor, harini – silk, tumbili – monkey) one at a time, each displayed for five seconds. In the testing phase, a Swahili word appeared, and participants had eight seconds to type in the Spanish word associated with it. Following Karpicke and Roediger (2008), who showed that repeated encoding and retrieval of information consolidates memories better, the paired associates test included three rounds of stimuli presentation and testing. After each stimuli presentation and prior to testing, participants were asked to solve 30 mathematical operations. Approximately 45 minutes after the last round, participants were asked to take a final recall test. For every correct word recalled, participants were given 1 point. Spelling or typographical errors were not taken into consideration.

Treatment task

Participants met online with the experimenter, a near-native speaker of English, and carried out a one-way communicative task which was video-recorded. The treatment materials are available at the IRIS database: https://doi.org/10.48316/S74gr-r3AQy. Two versions of the treatment task were created, each comprising 16 slides targeting the study's linguistic structure—eight for the -ing adjective context, eight for the -ed adjective context, and eight as distracters. The experimenter and the participant had two different sets of slides (see Appendix B and Appendix D). The experimenter’s slide had a question at the top and three possible answers: three pictures with their corresponding concept words underneath. The participants’ slide included a large picture and the same three small pictures with the same concept words as the experimenter (see Appendix C and Appendix E). One of the concept words was a noun derived from the target participial adjective that described the idea depicted by the large picture. During the treatment, the experimenter asked the prompt question at the top of her slide (e.g., “What does Brandon think about the spinach?”), and the participant had to answer the question by looking at the large picture on their slide and by using one of the concept words underneath the pictures (e.g., “Brandon thinks the spinach is disgusting”).

Feedback treatment

The feedback groups received oral CF at the end of the treatment task 1. The treatment was provided by playing an edited video recording of the treatment task with feedback instances inserted in the form of video overlays (see Appendix F). The editing of the video recordings was done on the spot while playing the original video to the participant with ScreenPal. This screencasting tool allowed the experimenter to record herself over the original video while stopping and adding explicit or implicit feedback each time the participant made a mistake on a target item. The no-feedback group completed the treatment task but did not receive any feedback during the replay.

For the explicit feedback group, the type of feedback provided was explicit correction, which first informed the learner about the inaccuracy of their utterance (i.e., “That is not correct”) and then introduced the reformulation of the error at the phrase level (i.e., “You should have said boring”). The implicit feedback group received explicit corrective feedback by means of recasts at the phrase level (i.e., “boring”) right after they made a mistake. Repairs after providing the correction were discouraged by swiftly asking participants to move on and continue watching the video.

The two feedback groups (explicit and implicit) were compared regarding the number of feedback instances they received via an independent sample t-test. The test compared the total number of feedback instances in the implicit group (\(N = 171, M = 9, SD = 3.88\)) and the explicit group (\(N = 114, M = 6.7, SD \))
and the results of the $t$-test indicated that the two groups were comparable ($t(34) = 1.93, p = .062$).

**Procedure**

Participants met with the experimenter online in one-on-one sessions via Skype on two consecutive days at approximately the same time (see Table 1 procedures). The experimenter administered the pretest and immediate posttest measures (GJT and OPT) on Days 1 and 2, respectively. Each time, the OPT was administered first and the GJT second. Google Slides was used for the OPT, and Google Documents was used for the GJT. The experimenter started the video conference, shared the links to the documents needed for the session, and asked the participant to share their screen. ScreenPal was used to record the sessions. The explicit associative memory test was administered on Day 1 before any language measures. The treatment task was also administered on Day 2 using Google Slides. All learners carried out the task with the experimenter. No feedback was provided during the task. At the end of the task, all participants watched a replay of the task. In the two feedback conditions, the same experimenter used ScreenPal to edit the recording while participants watched the replay. The experimenter stopped the replay right after each target error and inserted a feedback instance in the form of a video overlay.

**Table 1**

**Procedures**

<table>
<thead>
<tr>
<th>Day</th>
<th>Activity</th>
<th>Details</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Day 1</td>
<td>Cognitive Measure</td>
<td>Explicit associative memory test</td>
<td>20 mins (plus 45 mins wait)</td>
</tr>
<tr>
<td></td>
<td>Pretest</td>
<td>Oral production</td>
<td>30 mins</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Grammaticality judgment</td>
<td>20 mins</td>
</tr>
<tr>
<td>Day 2</td>
<td>Treatment</td>
<td>Participant-experimenter interaction</td>
<td>30 mins</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Replay of the interaction (the replay included CF for feedback groups only)</td>
<td>30 mins</td>
</tr>
<tr>
<td></td>
<td>Posttest</td>
<td>Oral production</td>
<td>30 mins</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Grammaticality judgment</td>
<td>20 mins</td>
</tr>
</tbody>
</table>

**Data analyses**

For the GJT, grammatical sentences that were accurately identified as correct received one point. Ungrammatical sentences that were labeled incorrect and properly corrected also received one point. When an ungrammatical sentence was labeled incorrect, but a feature other than the target structure was corrected, no credit was given. For each version of the GJT (A and B), reliability coefficients were computed using Cronbach’s alpha. The coefficients for the ungrammatical items were higher (version A = .91, version B = .83) than the coefficients for the grammatical ones (version A = .77, version B = .74). The coefficients for grammatical items were globally higher than .70, the minimum desirable alpha.
coefficient (Nunnally & Bernstein, 1994). However, not all the subversions with randomized orders reached this minimum of .70 in the case of grammatical items. The lower reliability of grammatical items, combined with the fact that the purpose of the GJT was to measure explicit knowledge, and that ungrammatical GJT items generally correlate strongly with measures of explicit knowledge (Ellis, 2005), led us to include only ungrammatical items in the analysis. An accuracy ratio score was calculated for the ungrammatical items by dividing the sum of accurate responses by 8, which is the total number of ungrammatical items.

Another accuracy ratio score was calculated for the OPT by dividing the number of accurate responses by 16, the total number of obligatory contexts created by the slides targeting the use of -ed and -ing adjectives. Responses containing an -ed adjective derived from the word presented on the slide were considered accurate if the item referred to an entity experiencing a feeling or state described by the adjective. Similarly, responses with an -ing adjective derived from the word on the slide were deemed accurate if the entity in the item was the causer of a feeling or state described by the adjective. When the participant created a context for the target structure and supplied the -ing or -ed forms, it was coded as accurate, even if the produced form was non-target like (e.g., “*satisfactioning” instead of “satisfying”). The experimenter coded the OPTs with a second rater coding 20% of the data to ensure reliability. The interrater reliability, calculated using simple percentage agreement, was 92.9%. Finally, all disagreements in scoring were discussed and resolved.

The following statistical procedures were used to determine the effects of feedback type and explicit associative memory. First, a multivariate analysis of variance was computed on relative gain scores (GJT and OPT as dependent variables) to establish whether there were any differences between the groups as an independent factor. Next, separate univariate tests were carried out to determine the effect of feedback condition on both outcome measures, followed up by post hoc tests for multiple comparisons using Tukey’s method. The magnitude of the gains in each group was also computed using Cohen’s d (Cohen, 1988). Finally, a regression analysis with interaction terms was performed for each outcome measure to determine whether the effect of explicit associative memory on learning gains differed between the groups.

Results

Before analyzing the data, we examined whether the variables in the study were normally distributed. Kolmogorov-Smirnov (K-S) tests yielded p values greater than .05, in a range between .21 and .97, and, therefore, normality was assumed.

The first research question in the study addressed the differential effect of explicit and implicit delayed feedback on learners’ linguistic development of English –ing and –ed participial adjectives as measured by learners’ ability to use the target structure in oral production (OPT) and to identify and correct target structure errors (GJT) successfully. To answer the research question, relative gain scores were first computed for each of the outcome measures. Relative gain is the difference between posttest and pretest scores divided by the amount to be learned (i.e., [posttest score - pretest score] / (1 - pretest score)), and it represents learners’ improvement as a proportion of their possible improvement (see Rogers, 2011). This transformation counteracts possible ceiling effects among learners who did not have as much room to grow, and it is preferable to other statistical methods that require a greater number of degrees of freedom to estimate parameters in order to compare improvement between groups (Rogers, 2011). Table 2 and Table 3 display the descriptive statistics for the pretest, posttest, and relative gain scores for the OPT and GJT, respectively. Figure 1 shows learners’ pretest to posttest development on each language measure (OPT and GJT).
Table 2

Descriptive Statistics for OPT Scores

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>M</th>
<th>SD</th>
<th>M</th>
<th>SD</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Feedback</td>
<td>16</td>
<td>0.64</td>
<td>0.20</td>
<td>0.57</td>
<td>0.28</td>
<td>-0.43</td>
<td>1.13</td>
</tr>
<tr>
<td>Implicit Delayed Feedback</td>
<td>19</td>
<td>0.43</td>
<td>0.19</td>
<td>0.58</td>
<td>0.21</td>
<td>0.21</td>
<td>0.36</td>
</tr>
<tr>
<td>Explicit Delayed Feedback</td>
<td>17</td>
<td>0.42</td>
<td>0.21</td>
<td>0.64</td>
<td>0.17</td>
<td>0.22</td>
<td>0.71</td>
</tr>
</tbody>
</table>

Note. Maximum Score = 1.0

Table 3

Descriptive Statistics for GJT Scores (Ungrammatical Items)

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>M</th>
<th>SD</th>
<th>M</th>
<th>SD</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Feedback</td>
<td>16</td>
<td>0.38</td>
<td>0.33</td>
<td>0.58</td>
<td>0.29</td>
<td>0.36</td>
<td>0.36</td>
</tr>
<tr>
<td>Implicit Delayed Feedback</td>
<td>19</td>
<td>0.19</td>
<td>0.21</td>
<td>0.47</td>
<td>0.39</td>
<td>0.39</td>
<td>0.44</td>
</tr>
<tr>
<td>Explicit Delayed Feedback</td>
<td>17</td>
<td>0.27</td>
<td>0.27</td>
<td>0.60</td>
<td>0.36</td>
<td>0.46</td>
<td>0.48</td>
</tr>
</tbody>
</table>

Note. Maximum score = 1.0

Figure 1

Pre-test Post-test Development
A multivariate analysis of variance (MANOVA) was performed on OPT and GJT relative gain scores with feedback condition (no-feedback vs. implicit delayed feedback vs. explicit delayed feedback) as the between-group factor. Box’s test of equality of covariance matrices was violated \((p < .05)\) and, therefore, Pillai’s Trace criterion was applied since it is argued to be the most robust statistic for general protection against departures from multivariate normality and homogeneity of variance-covariance (Tabachnick & Fidell, 2007). Using Pillai’s Trace, there was no group effect on the linear combination of the two dependent variables \((V = 0.16, F (4, 98) = 2.14, p = .081, \eta^2_p = .080)\). Separate univariate tests, however, revealed that feedback condition had a significant effect on the improvement of the groups’ performance on the OPT \((F (2, 49) = 4.025, p = .024, \eta^2_p = .14)\), but not on the GJT \((F (2, 49) = .247, p = .782, \eta^2_p = .01)\). Follow-up post hoc tests for multiple comparisons using Tukey’s method further showed that both the implicit delayed and explicit delayed feedback groups had improved significantly more than the no-feedback group \((M_{\text{diff}} = .67, 95\% \text{ CI } [.02, 1.31], p = .040, \text{ and } M_{\text{diff}} = -.67, 95\% \text{ CI } [.01, 1.33], p = .046, \text{ respectively}) and that they were not significantly different from each other \((M_{\text{diff}} = .001 \text{ and } 95\% \text{ CI } [.63, -.63], p = 1.000)\). There were no significant differences between the groups on the GJT \((p > .05)\).

Finally, Table 4 shows the magnitude of the differences \((d)\) between pretest and posttest scores. Although the \(d\) values confirm the results of the MANOVA analysis in general, they also provide additional information about the relative performance of the groups. As can be seen, the magnitude of the improvement that the two delayed CF groups showed on the GJT was medium.\(^2\) In addition, the CIs for these effect sizes did not cross 0, indicating statistically significant improvements for the CF groups from pretest to posttest\(^3\). The size of the improvement the no-feedback group showed was small and non-significant on the GJT. On the OPT, the size of the improvement for the explicit delayed feedback was medium but small for the implicit delayed feedback. The CIs show that only the explicit delayed, not the implicit delayed, group statistically significantly improved on the OPT. The no-feedback group’s performance went worse from pretest to posttest on the OPT, but the difference was not statistically significant.

Table 4

Summary of Pretest-Posttest Effect Sizes (\(d\))

<table>
<thead>
<tr>
<th>Group</th>
<th>GJT</th>
<th></th>
<th></th>
<th>OPT</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(d)</td>
<td>95% CI</td>
<td>(d)</td>
<td>95% CI</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Lower</td>
<td>Upper</td>
<td>Lower</td>
<td>Upper</td>
<td></td>
</tr>
<tr>
<td>No Feedback</td>
<td>0.62</td>
<td>-.13</td>
<td>1.37</td>
<td>-.031</td>
<td>-.65</td>
</tr>
<tr>
<td>Implicit Delayed</td>
<td>0.91</td>
<td>0.19</td>
<td>1.63</td>
<td>0.71</td>
<td>-.19</td>
</tr>
<tr>
<td>Explicit Delayed</td>
<td>1.07</td>
<td>0.30</td>
<td>1.90</td>
<td>1.01</td>
<td>0.04</td>
</tr>
</tbody>
</table>

In order to investigate whether explicit associative memory had any effects on learners’ linguistic development of participial adjectives depending on delayed feedback type, we first compared the average explicit associative memory scores in the three groups at the pretest stage in order to make sure that the groups were comparable (see Table 5 for a summary of the descriptive statistics).

Table 5

Descriptive Statistics for Explicit Associative Memory Test

<table>
<thead>
<tr>
<th>Group</th>
<th>Explicit associative memory</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(N)</td>
</tr>
<tr>
<td>No Feedback</td>
<td>15</td>
</tr>
</tbody>
</table>
A one-way analysis of variance (ANOVA) showed that there were no statistically significant differences between the groups at the outset of the experiment regarding their delayed recall scores on the explicit associative memory test ($F(2,47) = .223, p = .801$). To understand the moderating effect of explicit associative memory and to test for possible interactions with feedback condition (no-feedback vs. implicit delayed feedback vs. explicit delayed feedback), an analysis was performed using Interaction software (https://www.danielsoper.com/Interaction/, accessed on 5 November 2022). Two regression models were run, one for the OPT and one for the GJT. In the case of the OPT, there were no statistically significant results. The total model was not significant ($F(5, 40) = 1.693, p = 0.158, R^2_{\text{adjusted}} = .07$), and explicit associative memory was not a significant predictor ($\beta = -0.02, p = .401$). The two interaction terms for the delayed implicit and delayed explicit feedback groups (see Figure 2) were also statistically non-significant ($\beta = 0.022, p = .491$ and $\beta = 0.015, p = .669$, respectively).

**Figure 2**

Explicit Associative Memory and OPT Learning Gains

In the case of the GJT, the total model was not significant ($F(5, 40) = 1.228, p = 0.313, R^2_{\text{adjusted}} = .02$), and explicit associative memory was not a significant predictor ($\beta = -0.02, p = .148$). However, the two interaction terms comparing the slope of the baseline (no-feedback) group against the slopes of the delayed implicit and delayed explicit feedback groups were statistically significant in the case of the delayed implicit feedback group ($\beta = 0.032, p = .042$) and marginally significant in the case of the delayed explicit feedback group ($\beta = 0.034, p = .057$). As can be seen in Figure 3, the relationship between explicit associative memory and learning outcomes was positive in the two feedback groups (i.e., the higher explicit associative memory, the higher the learning gains), but it was negative in the no-feedback group (i.e., the higher explicit associative memory, the lower the learning gains).
Discussion

Does the explicitness of delayed feedback impact learners’ linguistic development differently in video-based CMC?

The first research question asked whether the explicitness of delayed feedback would have an impact on the acquisition of English -ing/-ed participial adjectives. Our results indicated that the two delayed feedback types were similarly effective when the groups’ gains were compared to each other. Their gains were also statistically higher than the no-feedback group’s gains on the OPT, but not statistically different from the no-feedback group’s gains on the GJT. Our complementary analysis including the comparison of the groups in terms of the size of their improvement from the pretest to the posttest and the confidence intervals of these effect sizes indicated that both treatment groups showed medium gains on the OPT, and these improvements were statistically significant. On the GJT, however, while the explicit delayed group showed a medium size improvement, the implicit delayed group showed a small size improvement, and only the explicit group showed a statistically significant improvement from the pretest to the posttest. Taking all this information into consideration, we argue that the evidence in favor of the effectiveness of explicit delayed feedback is slightly stronger than the evidence in favor of the effectiveness of implicit delayed feedback. This small difference might be the product of the unambiguous nature of the message provided in the explicit feedback condition. Perhaps the directness of the explicit feedback regarding the error and how to correct it communicated the negative evidence unambiguously and encouraged learners to link the errors they made with the correct linguistic information rather forcefully.

It may be helpful to contextualize our findings by comparing the effect sizes obtained in this study to the effect sizes in a recent feedback timing study (Canals et al., 2021), which used the same technological tool to provide delayed explicit CF on the same target structure. Although the studies reported almost identical effect sizes on the OPT (Canals et al., $d = 1.0$ vs. current study, $d = 1.01$), no similarity is observed when comparing the pretest-posttest effect sizes on the GJT (Canals et al., $d = 0.37$ vs. current study, $d = 1.07$). The discrepancy may be attributed to the difference in the operationalization of feedback.
delay. In the current study, feedback was provided at the end of the task, while in Canals et al. (2021), it was given 24 hours after the task. This variation in feedback delay might have impacted the quantity of knowledge measured by each task, as these tasks potentially tap into distinct types of knowledge.

It is argued that tasks such as untimed GJTs are more likely to tap into explicit knowledge (Ellis, 2005), which can be easily verbalized and brought into conscious awareness. Explicit knowledge is contrasted with implicit knowledge, which cannot be easily articulated or brought into conscious awareness (Ellis, 2005). According to Suzuki and DeKeyser (2017), only tasks that exhaustively prohibit conscious access to explicit knowledge (e.g., a word monitoring task) may measure implicit knowledge. Tasks such as some forms of OPTs that leave room for conscious access to knowledge should be considered measures of automatized explicit knowledge because knowledge may still be conscious even though it is accessed relatively rapidly and effortlessly. Based on Suzuki and DeKeyser (2017), the OPT and the GJT in this study may be considered to tap into automatized and nonautomatized explicit knowledge, respectively. Although the OPT was more meaning-focused than the GJT, some features of the OPT, such as the provision of a formally related word to the target adjectives, might have compromised its meaning-focused nature. In summary, a comparison of the results of the present study with those of Canals et al. (2021) suggests that a 24-hour delay may negatively impact nonautomatized explicit knowledge more so than automatized explicit knowledge.

Overall, our findings indicate that delayed CF, regardless of its explicitness, may be most beneficial for automatized explicit knowledge. That is, learners were able to proceduralize and automatize their explicit knowledge thanks to the practice opportunities the treatment task provided. This is in line with the transfer-appropriate processing principle (Lightbown, 2008), which would predict that video-based CF should be associated with higher oral language performance gains because the modality of the practice during which learners received CF matched the modality of assessment.

**Does explicit associative memory differentially affect learners’ linguistic development depending on the explicitness of delayed feedback in video-based CMC?**

The second research question asked whether explicit associative memory would differentially affect linguistic development depending on feedback type. The statistical analysis revealed that explicit associative memory was positively related to linguistic improvement on the GJT in the two delayed feedback groups, but negatively related to the performance of the no-feedback group. Considering a potential link between performance on the explicit associative memory test and intentional retention of information, this finding suggests that those learners with higher explicit associative memory ability might have been more able to make conscious links between delayed CF, the errors they had made during the treatment task, and, perhaps, prior nonautomatized explicit knowledge. The fact that the no-feedback group showed the opposite pattern can be interpreted as providing indirect evidence that the feedback treatment engaged a mental process that was facilitated by explicit associative memory. Unlike the participants who received the treatment, no feedback participants might have relied on their automatized knowledge to perform on both outcome measures, which could explain the negative relationship between their performance and explicit associative memory. To confirm this interpretation, however, future research should examine additional indicators of the explicitness of the cognitive processes involved, for example, via verbal reports.

One caveat to this interpretation is that not all the statistical evidence indicated improvement for the feedback groups in the GJT tasks. On the one hand, the feedback groups were not significantly different from the no-feedback group in the GJT. On the other hand, the effect size analysis showed that only the improvement of the feedback groups, not the no-feedback group, was statistically significant. Overall, while there is some evidence suggesting that delayed feedback led to improvement in the GJT task, this evidence is not as compelling as it could be.
Another issue that needs to be explained is the differential effect of explicit associative memory depending on whether the outcome measure was GJT or OPT. This finding could be due to the different types of knowledge these tasks measure, with the GJT tapping into non-automatized explicit knowledge and the OPT tapping into automatized explicit knowledge. It could be that participants’ prior non-automatized explicit knowledge about the target structure limited their improvement in this type of knowledge. Following this line of thought, the correlation between the limited gains made in non-automatized explicit knowledge and explicit associative memory might have been due to the fact that both the explicit associative memory test and the GJT involve decontextualized retrieval of previously encoded items/rules.

Conclusion, limitations, and pedagogical implications

The present study investigated the effects of the explicitness of delayed feedback on the acquisition of English -ing/-ed participial adjectives. Our data showed that the performance of the two delayed feedback groups was largely similar, and that the beneficial effects of delayed CF were more pronounced in the case of automatized explicit knowledge. We also detected a small difference in favor of explicit delayed feedback when effect sizes were considered. To confirm whether this small difference is real or an artifact of our small sample size, we encourage future researchers to replicate our study. In addition, we found evidence showing that processing under both implicit and explicit feedback conditions might be explicit based on the relationship between performance on a test of explicit memory and performance on a language measure that is more likely to gauge explicit knowledge (i.e., the GJT). This finding lent some support to Long’s (Granena & Yilmaz, 2022) prediction that delayed feedback leads to explicit processing regardless of feedback type.

The study has limitations affecting the generalizability of its results, including a small sample size and single-site design. In addition, the design did not include a delayed posttest and, as a result, we cannot determine the extent of knowledge retention in the long run. Also, our findings might not be generalizable to contexts that differ from the current study with respect to one or more of the following. The operationalization of feedback delay is one such factor. It is possible to provide CF with a longer delay, and this might lead to different results than ours. Other methodological parameters, such as the amount and type of feedback, proficiency level, and target language and structure, might have contributed to the results of our study. Finally, feedback provided using a different methodology may not produce the same results. To determine whether changes along these methodological parameters lead to similar results, more delayed feedback CF studies conceptually replicating the current study are needed.

It is important to highlight that the linguistic nature of our target structure might have contributed to the observed effectiveness of delayed CF. The facilitative effect of delayed feedback on the acquisition of English -ing and -ed found in this study coincides with the effect reported by Kim (2002). Although Kim (2002) and the present study differ in key methodological features such as type of feedback and feedback timing, both studies involved low intermediate learners and the same type of learning problem (i.e., a non-existent relevant semantic distinction in the L1). The fact that CF (regardless of type and timing) facilitated learning in both studies could be related to the linguistic nature of the target structure and to the fact that it involved a semantic feature that learners were able to notice through CF.

The pedagogical implications of this study for distance language classes are clear. Delayed CF, when provided in the manner described in this study, is a good investment, as it can result in greater linguistic improvement compared to carrying out communicative tasks without receiving any CF. Nevertheless, the type of CF has only a small impact on the effectiveness of CF. This means that educators facing time constraints to improve learners’ linguistic abilities may consider explicit delayed feedback, with the awareness that the advantages gained from this feedback type may be relatively minor. In contrast, those with more flexibility can choose either type of CF based on their preference. We recognize the practical
limitations of meeting with participants at the end of the task, and we recommend that this study be replicated with extended delays that can be more readily incorporated as an option in distance language education settings.

Acknowledgements

We would like to thank Diana Arroyo Ensuncho and Angela Donate Velasco for their invaluable help in this project.

Funding

This work was supported by a research grant by the Spanish Ministry of Science and Innovation in the Knowledge Generation program (#PID2021-128742NB-I00; PI: Gisela Granena).

Notes

1. This operationalization of delayed feedback as feedback provided immediately after task completion is the most common operationalization of delayed feedback in the feedback timing literature (e.g., Arroyo & Yilmaz, 2018; Li et al., 2016).

2. When pretest-posttest differences are considered, $d = 0.60$ represents a small effect, $d = 1.00$ a medium effect, and $d = 1.40$ a large effect according to Plonsky and Oswald (2014).

3. Confidence intervals provide an alternative approach to assessing significance. If the confidence interval does not contain zero, the $p$-value will be less than 0.05 (Larson-Hall, 2010).

4. It is important to stress that one cannot make strong claims about the nature of knowledge assessed by a particular task. Learners might have used both automatized and non-automatized explicit knowledge during these tasks.

References


Appendix A. Sample Item from the OPT for –ing Adjective

Compare A and B...

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>paintings by Picasso</td>
<td>architecture by Gaudí</td>
</tr>
</tbody>
</table>

...using a form of the word: fascination
Appendix B. Sample Item from the Treatment Task for –ed Adjective (Experimenter’s Slide)

How does Sally feel?

RELAXATION

STRESS

ANGER
Appendix C. Sample Item from the Treatment Task for –ed Adjective (Learner’s Slide)

Sally...

RELAXATION  STRESS  ANGER
Appendix D. Sample Item from the Treatment Task for –ing Adjective (Experimenter’s Slide)

What does Brandon think about the spinach and why?

DISGUST  DELICACY  HEALTH
Appendix E. Sample Item from the Treatment Task for –ing Adjective (Learner’s Slide)

The spinach...

DISGUST  DELICACY  HEALTH
Appendix F. Screen Capture of the Delayed-feedback Condition
About the Authors

Yucel Yilmaz is an Associate Professor of Second Language Studies at Indiana University. His research focuses on second language interaction and corrective feedback, computer-mediated communication, task-based language teaching, individual differences in second language acquisition, and explicit and implicit learning processes. Yucel Yilmaz is the corresponding author.

E-mail: yyilma@indiana.edu
ORCiD: https://orcid.org/0000-0001-5776-8342

Gisela Granena is an Associate Professor at the Universitat Oberta de Catalunya. She has published research on the role of cognitive aptitudes in both instructed and naturalistic learning contexts; aptitude-treatment interactions; task-based language teaching; measures of implicit and explicit knowledge; and the effects of early and late bilingualism on long-term L2 achievement.

E-mail: ggranena@uoc.edu
ORCiD: https://orcid.org/0000-0002-6322-9078

Laia Canals currently holds an Associate position at the Universitat Oberta de Catalunya. My research addresses technology-enhanced task-based language learning, the development of L2 through interaction in collaborative tasks in computer-mediated communication settings, the development of intercultural communicative competence, the development of digital skills, and online language teacher education.

E-mail: ecanalsf@uoc.edu
ORCiD: http://orcid.org/0000-0002-9605-012X

Aleksandra Malicka is currently Assistant Professor in the Arts and Humanities Department at the Universitat Oberta de Catalunya. She has done research on task-based language teaching, including task design, task sequencing and needs analysis.

E-mail: amalicka@uoc.edu
ORCiD: https://orcid.org/0000-0002-9966-5277