The Relative Effectiveness of Immediate and Delayed Corrective Feedback in Video-Based Computer-Mediated Communication

Abstract

Feedback timing is one of the factors that can moderate the effectiveness of feedback that has recently attracted the attention of corrective feedback researchers in SLA (e.g., Arroyo & Yilmaz, 2018; Henderson, 2021; Shintani & Aubrey, 2016). However, research to date has focused on either face-to-face or text-based computer-mediated communication. In this study, we investigated the relative effectiveness of immediate and delayed corrective feedback on the acquisition of -ing/-ed participial adjectives by English-as-a-foreign-language Spanish learners in video-based computer-mediated communication. Fifty-two participants took part in a communicative task in one of four groups (two experimental and two control). The immediate-feedback group received explicit corrective feedback during the task whereas the delayed-feedback group received the feedback 24 hours later by means of an edited video recording of the interaction. The effects of the feedback were measured through an oral production task and an untimed grammaticality judgment test. Although the results showed no differences between the two feedback timing groups and control participants on the grammaticality judgement test, both feedback groups outperformed the control group on the oral production task without showing statistically significant timing effects.

Keywords: feedback timing; computer-mediated communication; explicit correction; oral interaction; video-based feedback

INTRODUCTION

Corrective feedback (CF) understood as the response to learners' non-targetlike productions by their instructors or other learners has been established to have at least a facilitative role in second language (L2) acquisition in instructional settings (e.g., Doughty, 2001; Ellis, 1991; Gass & Mackey, 2006; Long, 1996; Pica, 1988). Several studies have tested the effectiveness of different types of corrective feedback (see Goo et al., 2015; Li, 2010; Lyster & Saito, 2010, for meta-analyses), but the focus of research has gradually shifted to the investigation of the factors that can moderate the effectiveness of feedback. One of these factors is feedback timing, or the time when CF is provided to a learner. Two types of feedback timing have been distinguished in the SLA literature: immediate and delayed. Immediate feedback is provided immediately after a learner's non-targetlike production. Delayed feedback is provided later, either at the end of the task, one day after the task, or several days after the task.

There is some evidence from CF studies investigating the acquisition of L2 grammar that indicates that immediate feedback is more effective than delayed feedback (e.g., Arroyo & Yilmaz, 2018; Li et al., 2016; Henderson, 2020; Shintani & Aubrey, 2016, but see also, Quinn, 2014). These studies, however, have investigated face-to-face oral CF (Li et al., 2016; Quinn, 2014) or written feedback, either web-based writing (Shintani & Aubrey, 2016), text-based computer-mediated communication (CMC; Arroyo & Yilmaz, 2018), or both face-to-face oral CF and text-based CMC (Henderson, 2020). Therefore, there is a need to investigate the role of feedback timing in other modalities, such as video-based CMC, which allow providing oral CF by means of video-comments by the instructor and which have important pedagogical implications for distance language learning environments where instructors cannot communicate synchronously with learners. In addition, previous studies

(e.g., Arroyo & Yilmaz, 2018; Li et al., 2016; Quinn, 2014) provided delayed feedback as error-correction pairs, which decontextualizes delayed feedback from the broader communicative context (i.e., the task), and on a list, one after the other, rather than spread out through interaction. These are features that have characterized delayed, but not immediate, feedback in previous research and, therefore, that could have been confounded with the variable under investigation (i.e., timing).

The present study set out to investigate the relative effectiveness of immediate and delayed CF on the acquisition of –ing/-ed participial adjectives by Spanish English-as-a-foreign-language (EFL) learners in video-based CMC. The use of –ing/-ed participial adjectives is problematic for Spanish learners of English because Spanish relies on a single adjective form and conveys the semantic differences between the –ing and –ed adjective forms using two different verb forms.. The modality investigated (i.e., video-based feedback) allowed us to address the methodological limitations of previous feedback timing research studies.

BACKGROUND LITERATURE

Feedback in Computer-Mediated Communication

CF can draw learners' attention to formal aspects of the target language and help them notice a mismatch between their production and the feedback received (Long, 1996). Theoretically, compelling cases have been put forward in favor of the necessity of providing the learner with "negative evidence," or with information about what is not possible in the L2, which can take the form of CF, among others (see Long, 2007). Empirically, there is ample evidence in support of the claim that CF facilitates language development (e.g., Goo & Mackey, 2013; Li, 2010; Russell & Spada, 2006; Yilmaz, 2012).

The provision of CF in face-to-face communication has been widely investigated in both classroom and experimental settings. This research has provided evidence of the effectiveness of CF in face-to-face communication (see Goo et al., 2015; Li, 2010; Long, 2007; Lyster & Ranta, 2013; Lyster & Saito, 2010; Lyster et al., 2013; Mackey & Goo, 2007; Nassaji, 2015; Nicholas, Lightbown, & Spada, 2001; Russell & Spada, 2006; Sheen & Ellis, 2011, for reviews and research syntheses).Several studies have shown that CF is also largely effective in CMC settings. CMC is an umbrella term that encompasses various forms of synchronous or asynchronous communication, as well as various modalities, text-based, audio-based, or video-based. Previous SLA research in CMC has been mostly conducted in text-based synchronous CMC or text chat (e.g., Baralt, 2013; Loewen & Erlam, 2006; Sauro, 2009; Yilmaz, 2012; Yilmaz & Yuksel, 2011), a modality that is considered similar to face-to-face communication, but with some potentially facilitative features in terms of L2 learning, such as visual saliency and increased processing time (Gonzalez-Lloret, 2014).

Baralt (2013) focused on task complexity and the effectiveness reformulating learners' ill-formed utterances (recasts), in face-to-face interaction and synchronous CMC of 84 college-level adult learners of Spanish as a foreign language. The results of two productive tasks and a multiple-choice test showed that both groups outperformed the control group, indicating that corrective feedback was effective regardless of communication mode in the acquisition of the Spanish past subjunctive. However, recasts in face-to-face interaction were more effective than in synchronous CMC if the task was cognitively more complex. In other experimental studies, synchronous CMC feedback was more effective than face-to-face feedback, as measured by recognition tests (Yilmaz, 2012) and oral production tasks (Yilmaz & Yuksel, 2011), and the effect did not change depending on feedback type, explicit correction or recasts (Yilmaz, 2012). There is also research showing that learning may not take place in a synchronous CMC context. Loewen and Erlam (2006) and Sauro (2009)

compared the effectiveness of providing recasts and metalinguistic information as CF in text-based synchronous CMC. Loewen and Erlam (2006) reported no statistically significant learning gains in response to either type of feedback in 31 elementary learners of English learning the regular Past tense, while Sauro (2009) found that only the metalinguistic-feedback group outperformed the control group in acceptability judgements (as pre, post and delayed posttests) in a study with 23 intermediate and advanced learners of English at a Swedish university.

Research on CF in CMC modalities other than text-based interaction is scarce in SLA. There are very few experimental studies which have looked at the provision of feedback in audio- or video-based synchronous CMC (Monteiro, 2014; Rassaei, 2017; Saito & Akiyama, 2017). Monteiro (2014) investigated the effectiveness of different feedback types in the development of implicit and explicit knowledge of the English past tense in video-based synchronous CMC with 42 Brazilian EFL learners of various levels. The study included two groups which received feedback (metalinguistic or recasts) and a control group. Treatment sessions were delivered through dyadic interactions using a video-conferencing tool. Implicit and explicit knowledge were respectively measured by a timed and an untimed grammaticality judgement test. The results of the study indicated no difference between the two feedback types which proved equally effective in helping learners develop explicit and implicit knowledge of the target structure.

In another study examining synchronous video-based interaction with native speakers, Saito and Akiyama (2017) examined the benefits of interactional CF (recasts and negotiation) on 30 college-level Japanese intermediate to advanced EFL learners' L2 oral abilities. L2 gains were measured by analyzing learners' oral performance on picture description tasks by two sets of raters (novice and experienced). The results showed that the learners in the experimental groups made significant pre- to post-test gains in comprehensibility, perceived

fluency (speech rate), lexical variation and grammatical accuracy (verb/article error ratio), whereas pronunciation did not significantly improve. The learners in the control group did not make any significant improvements in any of the areas. The findings provided support for the benefits of interactional feedback in a video-based synchronous CMC environment.

Finally, Rassaei (2017) tested the effectiveness of recasts on the use of definite versus indefinite articles by 57 Iranian EFL learners in two modes, face-to-face and video-based CMC. The effectiveness of feedback was measured by means of an oral production task and an error correction test, and was further examined in a stimulated recall interview. The results indicated positive effects of recasts in both communication modes without any significant differences between the experimental groups, but significant differences between the two experimental groups and the control.

The aforementioned studies suggest that CF in CMC settings can lead to an improvement in learners' L2 performance on various tests. However, research in CMC settings has been mostly conducted on text-based CF, while research on oral feedback in video-based modalities is scarce. In addition, to the best of our knowledge, there are no empirical studies investigating the effects of timing on video-based feedback in CMC.

Feedback Timing in L2 Learning

Several studies on feedback timing have shown advantages for immediate over delayed feedback in some outcome measures and communication modalities. In face-to-face communication, Li et al. (2016) examined the role of immediate and delayed feedback via prompts (i.e., feedback that pushes learners to repair their non-targetlike utterances) and recasts in 120 8th-grade Chinese EFL learners' acquisition of the English past passive. The immediate-feedback group showed an advantage over the delayed-feedback group in one of the outcome measures, a grammaticality judgment test measuring the development of explicit

knowledge, but not in the other measure, an elicited imitation task measuring the development of implicit knowledge. In addition, the immediate-feedback group showed larger effect sizes than the delayed-feedback group, which were sustained after two weeks. In another experimental study, Henderson (2020) investigated the effect of feedback timing on the acquisition of the Spanish subjunctive. The feedback administered in the study (i.e., repetition and reformulation of errors) was provided face-to-face or via text chat during or after (either immediate or delayed) two story-retell tasks. The immediate-feedback groups obtained greater pre-to-post gains in terms of noticing and linguistic improvement than the delayed-feedback groups regardless of the communication mode (face-to-face or text chat).

Other studies investigating the role of feedback timing in the effectiveness of oral CF have shown no difference between immediate and delayed feedback. For example, Quinn (2014) investigated the acquisition of the English passive construction among 19 intermediate-level adult ESL learners in an experimental study with two feedback conditions (immediate and delayed), combining prompts and reformulations. Three pretest and immediate/delayed posttest measures (an oral production test, an auditory grammaticality judgment test, and an error correction test) were administered before and after a treatment, which consisted of a 10-minute lesson on the target structure and three communicative tasks. The immediate-feedback group received feedback after each error and the delayed-feedback group at the end of each task. In the delayed-feedback group, learners were presented with the items that triggered their errors, prompted to try again, and then they received a reformulation regardless of their response. The results of the study showed no effects for either timing or feedback. In this case, and when comparing Quinn (2014) with the present study, one would need to be careful and account for the complexity of the target structure as a possible confounding factor which could moderate the effect of the feedback.

Arroyo and Yilmaz (2018) investigated the effects of feedback timing on errors that occurred during text-based synchronous CMC among 45 English L1 learners of Spanish ages 18 to 28. The authors looked at the acquisition of Spanish noun-adjective gender agreement in an experimental study where learner and experimenter interacted via text chat. Partial reformulations were provided as CF. These reformulations were administered by means of a chat feature in the immediate-feedback condition and by means of a written document compiling errors and reformulations in the delayed-feedback condition. The study found that the immediate-feedback group outperformed the delayed-feedback group on one of the measures (an oral production test), but not on a grammaticality judgement test. In another study, Shintani and Aubrey (2016) investigated timing effects using written feedback that was provided during or after web-based writing tasks on the acquisition of the hypothetical conditional among 68 intermediate-level Japanese college-level EFL learners. The feedback learners received involved reformulating errors using the comment box function of Google Docs and indicating the location of the errors on the original document. Both experimental groups outperformed the control group in the immediate posttest, but only the immediate-feedback group outperformed the control group in the delayed posttest.

Several of the empirical studies previously discussed provide evidence of an effect of CF on L2 development, with varying degrees of effects of feedback timing on different measures. In CMC contexts, such as in the present study, immediate feedback seems to be a better alternative than delayed feedback, as evidenced by the results in Arroyo and Yilmaz (2018) and Henderson (2020). Arroyo and Yilmaz argued that the effectiveness of immediate feedback may be related to differences in the cognitive mechanisms involved in the processing of CF under different timing conditions. Following theoretical claims made in the focus on form (FonF) literature (e.g., Doughty, 2001), the short time lapse between feedback and error in immediate feedback may help learners make the mental comparison between

communicative intent, output, and the input conveyed through feedback while they are still held in working memory. As the distance between the error and the feedback increases, the effectiveness of the feedback may tend to diminish (Doughty & Long, 2003).

A key feature that characterized delayed feedback as administered in the studies previously reviewed is that it was decontextualized from the meaning-based activity where the errors took place and was presented as part of a list of error-correction pairs or episodes. Therefore, timing, the variable under investigation, could have been confounded with other methodological features of the delayed feedback treatment, an issue to which we turn next.

Methodological Limitations in Feedback Timing Research

Previous research on feedback timing in SLA has some methodological limitations. First, in immediate-feedback conditions, feedback instances are naturally spread throughout the interaction, while, in delayed-feedback conditions, feedback instances were typically provided all at the same time without any intervening interactional elements (e.g., Li et al. 2015; Arroyo & Yilmaz, 2018). As a result, feedback conditions in previous research studies were not only different in terms of feedback timing, but also in terms of feedback spacing. A second related methodological issue is that delayed feedback in previous research was provided either through a written list of errors with corrections (e.g., Arroyo & Yilmaz, 2018) or by repeating the learner's error before providing a correction (e.g., Li et al., 2016). The assumption made in the literature seems to be that learners need to remember having made the error that is re-presented to them at the delayed-feedback stage (Arroyo & Yilmaz, 2018). However, this way of providing delayed feedback decontextualizes feedback instances from the broader communicative context (i.e., the task) where errors were made. As a result, a decontextualized delayed feedback instance might make it harder for learners to recall the memory of having made the error. Recalling personal experiences, such as committing an

error during a communicative task, belongs to the domain of episodic memory (Tulving, 1972), which involves "recalling the what happened, where, and when of events" (Wheeler & Gabbert, 2017, p. 2). According to the encoding specificity principle of memory (Tulving & Thomson, 1973), an effective retrieval cue should be compatible with the memory trace created during encoding and the match between the cue and target memory should be high (Wheeler & Gabbert, 2017). Contextual cues can enhance the completeness of recall by spreading activation to information that was not initially accessible (Smith et al., 2014). Under the delayed-feedback conditions used in previous research, the cue provided (oral or written error-feedback pairs) may have been insufficient for learners to recall the memory of having made the error, and may have been partially responsible for the relative ineffectiveness of delayed feedback. Overall, although the ways in which delayed feedback was provided in previous research can be considered pedagogically realistic (i.e., ecologically valid), they may have interfered with determining the true potential of delayed feedback.

THE PRESENT STUDY

There is some evidence from CF studies (Arroyo & Yilmaz, 2017, 2018; Li et al., 2016) suggesting that immediate feedback is more effective than delayed feedback. However, these studies have investigated face-to-face communication (Li et al., 2016) or text-based communication, either web-based writing (Shintani & Aubrey, 2016), text-based CMC (Arroyo & Yilmaz, 2017) or both face-to-face and text-based CMC (Henderson, 2020). Therefore, there is a need to investigate the timing of CF in other modalities such as video-based CMC which has been shown to facilitate the effectiveness of feedback (Monteiro, 2014; Rassaei, 2017; Saito & Akiyama, 2017). From a pedagogical perspective, the video-based modality may be particularly relevant for distance language learning environments where there is no synchronous interaction between learners and instructors

given that video-based feedback can be synchronous or asynchronous in nature. An asynchronous type of instruction occurs delayed in time, and it is one of the most prevalent forms of online education, on its own or in combination with a synchronous online modality (e.g., Johnson & Aragon, 2003; see Canals et al., 2020, for an example in L2 learning). When teacher-learner communication is asynchronous and learners video-record themselves speaking or interacting with other learners, the provision of synchronous oral CF by the instructor is not feasible. While asynchronous video-based feedback could be of use because it is delayed in nature, the relative effectiveness of immediate and delayed feedback in this modality needs to be empirically investigated.

In addition, investigating feedback timing in video-based CMC allows addressing certain methodological limitations of previous feedback timing research in SLA. Video-based CMC allows using video editing tools to add delayed feedback instances as video overlays, that is, as video annotations by the instructors/researchers to the recorded learner-learner interactions. The learner can then view a playback of this edited video-based CMC interaction. When delayed feedback is provided this way, feedback instances are spaced in the interaction. In addition, the broader context of the interaction where the errors took place is preserved. By editing and playing the video-based CMC interaction back to the learners in the delayed-feedback condition, the present study aimed at improving the internal validity of previous research on the effects of feedback timing. Finally, the current study aimed to extend previous research on CF timing by choosing a feedback type that has not been investigated in the context of CF timing research. Previous literature has shown that explicit corrections lead to high learning effects (Li, 2010), and that they are more effective than some other feedback types, such as recasts (Yilmaz, 2012, 2013). Considering that language teachers may show an interest in using explicit corrections due to their effectiveness, a new study on feedback timing using explicit corrections is warranted.

The following research question guided our research:

Does feedback timing in video-based CMC affect the development of English -ing/-ed participial adjectives as measured by (a) an untimed grammaticality judgment test and (b) an oral production task?

METHOD

Study Design

The study followed a pretest-posttest-control group design (see Table 1). Participants were randomly assigned to two treatment groups, immediate feedback and delayed feedback. The immediate-feedback group received oral CF as they carried out the treatment task. The delayed-feedback group received oral CF 24 hours after carrying out the treatment task by means of a playback of the treatment task with feedback instances added as video overlays. A 24-hour delay was selected for methodological purposes to be able to fill a gap in the previous literature, where delayed feedback has been mostly provided immediately after the treatment task (e.g., Arroyo & Yilmaz, 2018; Li et al., 2016). It was also considered a more ecologically valid decision considering the prevalence of pedagogical contexts, such as distance learning environments, where instructors' feedback cannot be delivered immediately after task completion. A delayed posttest was not included because participants were attending language classes at the time of the study and, therefore, outside-the-experiment exposure to the target structure posed a threat to experimental validity.

To keep the time between the provision of feedback and the pretest/posttest measures constant in the two treatment groups, the immediate-feedback group was administered the treatment task (with feedback) on Day 2, while the delayed-feedback group was administered the treatment task (with no feedback) on Day 1 and the playback of the treatment task (with feedback) on Day 2 (see Table 1). This created an imbalance between the conditions. While

errors in the immediate-feedback group were corrected during the treatment task on Day 2, errors in the delayed-feedback group were ignored on Day 1 and corrected at the playback stage on Day 2. As a result, a decision was made to also administer a playback of the treatment task with all feedback instances removed to the immediate-feedback group on Day 3 to control for any potential confounds.

Table 1

Study Design

Group	**DAY 1**	**DAY 2**	**DAY 3**
Immediate (n=16)	Pretest	Treatment task (feedback)	Playback of treatment task (no feedback)+Posttest
Delayed (n=16)	Pretest+Treatment task (no feedback)	Playback of treatment task (feedback)	Posttest
Control 1 (n=10)	Pretest	Treatment task	Playback of treatment task+Posttest
Control 2 (n=10)	Pretest+Treatment task	Playback of treatment task	Posttest

Two control groups were used to control for different sources of confounding. Control 1 participants were administered the playback of the treatment task on Day 3, whereas control 2 participants were administered the playback of the treatment task on Day 2. This allowed controlling for the time difference between video playback and pretest/posttest in the two treatment groups. These two control groups were combined into a single control group (n = 20) to gain greater statistical power, after checking that there were no statistically significant differences between them in their pretest scores and in their level of improvement from pretest to posttest (see Appendix A).

Participants

The participants in the study (N = 52) were Spanish learners of English as a foreign language at a Spanish university taking a pre-intermediate course (i.e., B1, according to the Common European Framework for Reference of Languages). There were 17 males (33%) and 35 females (67%) and their average age was 34.6 (SD = 8.3). The study was advertised in 15 different classrooms as an opportunity to practice conversation with a native speaker. Even though the researchers had access to a relatively large number of potential participants, the drop-out rate turned out to be quite high due to the experimental nature of the study. The participants had to meet online with the researcher on three consecutive days at the same time which had a detrimental effect on participant recruitment and on the number of participants that completed all the sessions due to two main reasons. First, since there was no academic or financial incentive for participants to connect other than to practice their English, some participants connected on the first day, but not on the second or the third one. Second, personal and professional commitments prevented participants from connecting at roughly the same time on three consecutive days. In both scenarios, the participants needed to be discarded.

Target Structure

The target linguistic structure in this study were English –ing and –ed participial adjectives. These are adjectives that are derived from transitive verbs of state and emotion and that are typically presented in grammar books and textbooks in pairs such as interesting/interested, boring/bored, or surprising/surprised. For example, boring and bored are formed from the present and past participle of the verb bore, respectively. While boring in "The boy is boring" describes what the boy is, bored in "The boy is bored" describes what happened to the boy. The use of the–ing and –ed participial adjectives is problematic for English learners of all

levels across a wide variety of first language groups (Scovel, 1974). In the case of Spanish learners of English, the acquisition of –ing/-ed adjectives poses a learning problem because Spanish indicates the differences in meaning between –ing and –ed adjectives using two different verb forms, both translated as to be in English (i.e., ser and estar), either with the same participial adjective, the past participle (e.g., aburrido for both bored and boring), or with two different forms of the past participle (e.g., confundido vs. confuso for confused and confusing, respectively). A common mistake for Spanish learners of English is to say "I'm boring" (i.e., 'Soy aburrido') when they actually mean "I'm bored" (i.e., 'Estoy aburrido').

A survey sent out to the English language instructors at the institution where the study was conducted asking instructors to list any observed learning problems confirmed that -ing/-ed adjectives were a persistent problem for learners at the B1 proficiency level. We determined the list of participial adjectives derived from emotive verbs to be included in the study by consulting grammar reference books such as Larsen-Freeman and Celce-Murcia (2015). We then refined the list by considering the depictability of these adjectives and/or the predictability of the meaning of the adjectives based on Spanish.

Treatment

Treatment Task. Each learner carried out a one-way communicative task with a native speaker of English (hereafter, the experimenter). The task was carried out online using a video-conferencing tool (Skype) and recorded using a screen recording and video editing tool (Screencast-O-Matic). Each task included 24 slides: 16 slides targeting the linguistic structure of the study (8 –ing and 8 –ed adjective contexts) and 8 distracters. The experimenter saw a question at the top of their slide and three different pictures paired with three concept words underneath, each of which could potentially answer the question, while the participant saw a large picture at the top and the same three picture-concept word pairs the experimenter saw

underneath(see Appendices B, C, D, and E). The concept words were a noun form that derived from the same emotive verb as the adjective. The two remaining concept words were distractors. The purpose of including a distractor concept word was to focus learners' attention on the target word by determining the best concept that matches the scene displayed in the item. The meanings of these words were made transparent to the learners through the use of small pictures that were paired with them. The experimenter asked the question on their slide (e.g., 'How does Sally feel?') and the learner's task consisted in answering the question by describing the large picture on their slide with the help of one of the three pictures with captions located underneath (e.g., 'Sally feels relaxed'). Task instructions were provided in the learners' L1. During the administration of the OPT, the experimenter made sure that the participant selected the correct concept word and used that concept to describe the picture.

Feedback Treatment. All learners carried out the treatment task with the experimenter, but only the immediate and delayed feedback groups received oral CF on their errors. The type of feedback provided was explicit feedback which was selected because previous research has shown that it is more effective than other forms of feedback (Li, 2010; Yilmaz, 2012). Specifically, the type of explicit feedback provided was explicit correction. Explicit correction directly informs the learner about the accuracy of their utterance (i.e., 'that's not correct') and explicitly introduces the reformulation of the error (see Example 1) at the phrase level focusing on the phrase which contained the non-targetlike form.

Example 1

Experimenter: How is the spinach in Brandon's opinion?

Learner: *The spinach is very disgusted.

Experimenter: That's not correct; you should say 'disgusting.' Next.

Immediately after feedback was provided, learners were prompted to move on to the following item to prevent them from repairing, repeating, reformulating or producing any type of modified output.

During the treatment, the difference between the immediate- and delayed-feedback groups was the timing of the CF they received on non-targetlike forms. The learners in the immediate-feedback group received oral CF while performing the treatment task during video-conferencing, whereas the learners in the delayed-feedback group received oral CF 24 hours after performing the task. To be able to isolate the timing variable from other confounding factors, the immediate and delayed-feedback conditions were kept as similar as possible. Video editing was used to control for differences between the groups. To this end, CF to the delayed-feedback group was provided by means of a playback of the entire treatment task after adding oral feedback instances as video overlays immediately following learners' non-targetlike forms (see Appendices F and G). In addition, the immediate-feedback group was asked to view the entire treatment task they had carried out for a second time after all feedback instances had been removed. Therefore, both groups viewed the interaction for a second time, the immediate group with feedback instances removed and the delayed group with feedback instances added. The participants in either group did not have the control over the playback feature of the video of their performance which was controlled by the researcher at all times so that they could not rewind, pause or rewatch the video more than once.

The two treatment groups (immediate and delayed) were compared regarding the number of feedback instances they received via an independent sample *t*-test. Group means were 14.81 (SD = 6.96) and 15.93 (SD = 8.77), respectively. The results of the test showed that there were no significant differences between them (p = .957).

Instruments

A grammaticality judgement test (GJT) and an oral production task (OPT) were used as outcome measures for the pretest and posttest. For each measure, two versions including different items were created and counterbalanced across participants within each treatment condition at each time (pretest and posttest). Each version had three different randomized orders to control for possible order effects.

In the OPT, learners were shown sets of two items on a slide presentation and were asked to compare them using a form of the word that was displayed at the bottom of the screen (see Appendix H). This task aimed at measuring learners' knowledge of the target structure that could be deployed during communicative use when their attention was primarily focused on meaning. The OPT included 16 contexts for the use of the target form for a total of 16 target items and 8 distractors. The GJT was used to measure learners' knowledge of the target structure when their attention was primarily focused on the correctness of the linguistic form. In this task, learners were given a sheet including a list of sentences and they were 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 16 target items and 8 distractors, half of them grammatical and half ungrammatical. There were no overlaps between the treatment, GJTs, and OPTs in terms of target items.

Procedure

Learners met with the experimenter online in one-on-one video-conferencing sessions via Skype on three consecutive days (Day 1, Day 2, and Day 3). On the day of the session, the experimenter, a native speaker of English with language teaching experience, called the

participant, shared the links to the documents needed for the session and asked them to share their screen. A screen recorder (Screencast-O-Matic) was used to record the sessions.

The experimenter administered the pretest and posttest measures (GJT and OPT) on Day 1 and 3, respectively. At each time, first the OPT and then the GJT were administered. Google Slides was used for the OPT and Google Documents for the GJT. It took approximately 25 and 15 minutes for the learners to complete the OPT and the GJT, respectively.

The treatment task was also administered using Google Slides and it took between 15 and 20 minutes. The same experimenter edited each video recording of the treatment task using Screencast-O-Matic. If the recording belonged to a participant in the delayed-feedback group, the experimenter added feedback instances by pausing the recording, recording herself while providing the feedback, and adding the recording as a video overlay. If the recording belonged to a participant in the immediate-feedback group, the experimenter removed all the feedback instances.

Data Analysis

Grammatical sentences in the GJT that learners 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. Two accuracy scores, one for grammatical and one for ungrammatical items, were calculated by dividing the sum of accurate responses by the total number of items in that category. Reliability coefficients were computed for the grammatical and ungrammatical items in each version of the GJT (A and B) using Cronbach's alpha. The coefficients for the grammatical items were low, indicating that these items did not discriminate well (version A = .50, version B = .39). The descriptive data

of these items showed a ceiling effect, which means that participants' response accuracy was very high. The coefficients for the ungrammatical items were good (version A = .91, version B = .88). Therefore, the grammatical item scores were not submitted to statistical analyses.

In the case of the OPT, learners' video-recorded oral responses were transcribed and coded. An accuracy score was computed by dividing the number of accurate responses by the number of obligatory contexts created. A response was coded as accurate when the participant created a context for the target structure and supplied the –ing or –ed ending accurately, even if the form supplied was not targetlike (e.g., *temptating). Coding reliability was established by having two native English-speaking coders listen to and code 20% of the data. The percentage agreement between the two coders was 96.7%. Disagreements in scoring were discussed and resolved.

The following statistical procedures were followed to determine the effect of feedback timing on each outcome measure (GJT and OPT). First, a one-way ANOVA was computed on pretest scores to establish whether there were any differences between the groups. Next, a mixed-design ANOVA was carried out with group as a between-subjects factor and time as a within-subjects factor. If the analysis showed a statistically significant Group-by-Time interaction, we followed up by comparing each group's pretest and posttest scores via paired-samples *t*-tests and by comparing the groups' posttest scores while controlling for pretest scores via ANCOVA. We also calculated gain scores for each outcome measure by subtracting the groups' posttest scores from their pretest scores and by computing the magnitude of each group's gains per outcome measure using Cohen's *d* (gain scores/pooled *SD*) as an index of effect size.

RESULTS

Before analyzing the data, we inspected the kurtosis and skewness values of all scores, using the range between −2 and +2 as the target range (George & Mallery, 2010) to determine whether data conformed to normality. All the scores fell within the −2 and +2 range with values between −0.80 and +0.98, and therefore, normality was assumed.

Grammaticality Judgment Test

Table 2 shows the descriptive statistics for GJT scores using ungrammatical items (see Appendix I for the descriptive statistics for grammatical items). As can be seen, the standard deviations relative to the means for most of the pretest, posttest and gain scores were quite large, which indicates a wide spread in the scores and, therefore, considerable variation in the extent to which participants were able not only to judge the grammaticality of the items but also to accurately identify the source of the ungrammaticality by correcting the error.

Before conducting the mixed-design ANOVA to determine whether the level of improvement from pretest to posttest changed depending on group, a univariate ANOVA was carried out on the groups' pretest scores. The results of this analysis showed that the groups were comparable in their prior knowledge of the target structure, $F(2, 49) = .94$, $p = .399$, $\eta^2 = .037$. The mixed-design ANOVA revealed that Group, $F(2, 49) = .72$, $p = .490$, $\eta^2 = .03$, and the interaction between Time and Group, $F(2, 49) = 1.59$, $p = .215$, $\eta^2 = .05$, were not statistically significant. However, Time was statistically significant, $F(1, 49) = 15.78$, $p < .001$, $\eta^2 = .23$, suggesting that all groups improved from pretest to posttest, but that the extent to which each group improved was comparable.

Table 2

Descriptive Statistics for Grammaticality Judgment Test Scores

Group		Pretest		Posttest		Gains	
	N	*M*	*SD*	*M*	*SD*	*M*	*SD*
Control	20	0.21	0.25	0.34	0.32	0.13	0.27
Immediate Feedback	16	0.19	0.21	0.48	0.28	0.29	0.25

Delayed Feedback	16	0.31	0.30	0.42	0.30	0.11	0.41

Oral Production Task

The descriptive statistics for the OPT scores appear in Table 3. The table shows that the pretest-posttest gains of the immediate and delayed feedback groups were much higher than the gains of the control group. The two feedback groups' gains, however, were similar. The standard deviations in the OPT were not as high as in the GJT, indicating that the range of the scores in the GJT was not as wide. Before we carried out the mixed-design ANOVA, we first compared the performance of the three groups on the pretest and found that the three groups' pretest scores were not statistically significantly different, $F(1, 49) = .90$, $p = .413$, $\eta^2 = .035$. The mixed-design ANOVA revealed that Time, $F(1, 49) = 38.15$, $p < .001$, $\eta^2 = .48$, and the interaction between Time and Group, $F(2, 49) = 3.38$, $p = .042$, $\eta^2 = .10$, were statistically significant. However, Group, $F(2, 49) = 2.56$, $p = .087$, $\eta^2 = .07$, was not statistically significant. To probe the interaction between Time and Group, we ran paired-samples *t*-tests comparing pretest and posttest scores in each of the groups. These analyses revealed that all the groups improved statistically significantly from pretest to posttest: control group, $t(19) = -2.12$, $p = .047$; immediate group; $t(15) = -4.52$, $p < .001$; delayed group, $t(15) = -3.47$, $p = .003$. In addition, we compared the groups' posttest scores after controlling for their pretests scores. The results of this ANCOVA showed a significant Group effect, $F(2, 48) = 5.07$, $p = .010$, $\eta^2 = .13$, and a significant covariate effect for pretest scores $F(1, 48) = 22.16$, $p < .001$, $\eta^2 = .28$. Bonferroni-adjusted pairwise comparisons further revealed significant differences between the immediate feedback and the combined control group ($MD = -.16$, $p = .024$), as well as between the delayed feedback and the combined control group ($MD = -.15$, $p = .035$). There were no differences between the immediate feedback and the delayed-feedback groups ($MD < .01$, $p = 1.000$).

Table 3

Descriptive Statistics for Oral Production Task Scores

Group		Pretest		Posttest		Gains	
	N	*M*	*SD*	*M*	*SD*	*M*	*SD*
Control	20	0.37	0.22	0.45	0.20	0.08	0.16
Immediate Feedback	16	0.36	0.22	0.60	0.23	0.24	0.21
Delayed Feedback	16	0.45	0.20	0.64	0.18	0.19	0.22

To summarize, two sets of analyses were conducted, one on the GJT scores and another on the OPT scores. In the case of the GJT, neither the immediate-feedback group nor the delayed-feedback group outperformed the control group. There were no statistically significant differences between the two feedback groups either. Despite the lack of statistical differences between the groups, we should note that the magnitude of the gains in the immediate-feedback group ($d = 1.17$) was larger than the magnitude of the gains in the other groups ($d = 0.37$ in the delayed-feedback group and $d = 0.45$ in the control group; see Table 4 for effect sizes).

In the case of the OPT, both feedback groups statistically outperformed the control group. However, there were no statistically significant differences between the two feedback groups. This finding was further supported by a comparison of the magnitude of the gains of the two feedback groups from pretest to posttest. While the magnitude of the gains in the OPT in both groups was large, the difference between the two groups in terms of magnitude was negligible ($d = 1.07$ and $d = 1.00$; see Table 4).

Table 4

Effect Sizes

Group	GJT			OPT		
	d	95% CI		*d*	95% CI	
		lower	upper		lower	upper
Control	0.45	−0.18	1.07	0.38	−0.25	1.00
Immediate	1.17	0.39	1.89	1.07	0.30	1.78
Delayed	0.37	−0.34	1.06	1.00	0.24	1.71

DISCUSSION

This study set out to investigate the effect of timing on the effectiveness of CF in a CMC environment. The research question that guided the study focused on the extent to which feedback timing would affect L2 learners' development of English -ing/-ed participial adjectives in video-based CMC. To answer this question, we first analyzed GJT scores. The results of a mixed-design ANOVA comparing the three groups (i.e., immediate feedback, delayed feedback, and control) showed that the three groups improved significantly from pretest to posttest, but that there were no significant differences between each group's improvement.

Although there were no statistically significant differences between the groups, a difference could be observed in the magnitude of the improvement in the groups. The gain for the immediate-feedback group was 0.29 and the gain for the delayed-feedback group was 0.11. Following Plonsky and Oswald's (2014) criteria, where $0.40 \leq d < 0.70$ implies a small effect, $0.70 \leq d < 1.00$ a medium effect, and $d \geq 1.00$ a large effect, the magnitude of the gains on the GJT in the immediate-feedback group was large ($d = 1.17$), whereas in both the delayed feedback and the control group the magnitude was small ($d = 0.37$ and $d = 0.45$, respectively). The superiority of the effect size observed in the immediate-feedback group in this study coincides with the superiority of the effect size of immediate-feedback conditions observed in previous feedback timing studies. Li et al. (2016) did not find significant differences between the effects of immediate and delayed feedback on the GJT, but the effect sizes of immediate feedback were larger than those of delayed feedback in 80% of the pairwise contrasts when the two feedback types were compared with each other. Similarly, Shintani and Aubrey (2016) did not find statistically significant differences between the immediate and delayed-feedback conditions but effect sizes for the posttests were larger for immediate feedback. Finally, in Arroyo and Yilmaz (2018), the effect sizes of the gains on the

two outcome measures in the immediate-feedback group were also superior on average than in the delayed-feedback group.

The consistently larger effect size shown by immediate feedback-conditions is a pattern that suggests that immediate feedback may be more effective in improving L2 learners' accuracy than delayed feedback and helps shed some light on non-statistically significant differences between experimental conditions or between experimental conditions and control groups in feedback timing studies. While a look at the statistical *p* values in our results would lead us to conclude that there was no effect for timing as measured by gains on the GJT, a look at the effect sizes indicates that at least immediate feedback seems to have assisted acquisition to a larger extent ($d = 1.17$) than delayed feedback ($d = 0.37$) or no feedback ($d = 0.45$).

Unlike GJT scores, the analysis of OPT scores revealed that both feedback groups, immediate and delayed, outperformed the control group, but with no statistically significant differences between the effects of the two types of feedback. In other words, L2 gains on the OPT were comparable regardless of whether learners received immediate or delayed feedback (0.24 and 0.19, respectively). The magnitude of the effect was large in both cases ($d \geq 1.00$) and the difference negligible ($d = 1.07$ and $d = 1.00$, respectively). The finding that immediate and delayed feedback can be equally effective in L2 learning and have an effect of equal magnitude contrasts with previous results in the SLA literature, which either showed an advantage for immediate feedback (e.g., Arroyo & Yilmaz, 2018; Henderson, 2020; Li et al., 2016; Shintani & Aubrey, 2016) or no effects for either type of feedback (e.g., Henderson, 2021; Quinn, 2014). A factor that could account for the mixed findings is the modality investigated and how this modality affected the way delayed feedback was provided. Previous research on feedback timing was conducted either in face-to-face communication (Li et al., 2016; Quinn, 2014) or in the written mode, via text chat (Arroyo & Yilmaz, 2018;

Henderson, 2021), web-based writing (Shintani & Aubrey, 2016), or in both modes (Henderson, 2020). In the present study, it was oral, like in face-to-face communication, but video-based, which allowed editing the video recording of the interaction to insert video-based feedback instances immediately after learners' non-targetlike productions. The resulting edited video recording of the interaction, which was played back to the learners, provided them with feedback that was delayed in time but, unlike in previous studies, that was contextualized, that is, delivered in the context of the interaction where the errors took place. Under these conditions, feedback instances are spaced, or naturally spread, throughout the interaction.

This way of providing delayed feedback by conveying needed information about the target language in context rather than as part of a decontextualized list of errors and corrections (e.g., Arroyo & Yilmaz, 2018) is more aligned with theoretical claims in the FonF literature about the facilitative role of in L2 learning. The provision of feedback has been viewed as one of the major ways of implementing FonF. FonF "involves reactive use of a wide variety of pedagogic procedures to draw learners' attention to linguistic problems in context, as they arise during communication" (Long, 2015, p. 317). This is considered to facilitate the processing of form and form-function mapping because the learner already has prior comprehension of at least part of the message. In addition, the playback of the interaction might have helped learners activate the memory of having made the error. Arroyo and Yilmaz (2018) suggested that one of the reasons why delayed feedback did not show any effectiveness in their study could have been that learners were not able to link the errors that were re-presented to them on a computer screen in a decontextualized way to the errors they made during their task performance. In other words, seeing their own errors and the corrections was not enough for the learners to benefit from feedback. If, as argued in the cognitive psychology literature, the provision of contextual cues can enhance the recall of a

past event and help activate memory traces (Smith et al., 2014), the provision of delayed feedback in the context of the interaction where the errors took place may help promote the comparison of communicative intention, output, and input that is considered to facilitate learning in the case of other types of feedback, such as implicit immediate reformulations, or recasts (Doughty, 2001).

The importance of contextual information for effective memory retrieval is defended by the encoding specificity principle of memory (Tulving & Thomson, 1973). According to this theoretical framework, the recall of personally experienced events, which falls within the domain of episodic memory, is improved when information available at encoding (i.e., surrounding cues) is also available at retrieval. In other words, memory is better when the encoding and the retrieval context match. In delayed-feedback conditions, learners are expected to retrieve a past experience (i.e., the interaction) where they made an error. They must go back in time to a given, specific episode. The provision of contextualized delayed feedback, as opposed to isolated error-correction pairs, might have helped learners recall the personal experience of having committed an error during a communicative task (i.e., the what happened, where, and when). Successful recall of this information from memory may have contributed to the effectiveness of delayed feedback found in the present study.

However, this does not explain why the effect of delayed, and also immediate, feedback was observed on the OPT but not on the GJT.One possibility is that the learners in this study already had prior explicit knowledge of the target structure and that the treatment task contributed to automatize that knowledge for use in oral production. According to skill-learning theory, the declarative representation of a linguistic feature is first proceduralized through mechanical practice and eventually automatized through production practice under real operating conditions (DeKeyser, 2007). Both the treatment task and the OPT involved comparing and describing items and learners not only carried out but also

viewed a video recording of the treatment task, either with or without feedback instances, depending on the timing condition. All these factors could have contributed to the automatization of explicit knowledge of the target structure and its retrieval in the OPT[1].

One difference, however, between immediate and delayed feedback from the point of view of the theoretical claims made in the FonF literature is that, when learners receive immediate feedback, they are vested in the communicative exchange (Long, 2007), whereas in a delayed-feedback condition learners may be attending, and even motivated, but they are not actively participating in the communicative exchange nor using the L2 to communicate. In our delayed-feedback condition, feedback processing was contextualized, but the processing occurred outside the communicative event in which the errors were committed. This could be a potential psycholinguistic advantage of immediate feedback that could contribute to explain the higher effect sizes associated with immediate feedback. However, further research is needed to shed more light on how feedback is processed under different feedback timing conditions and modalities. The psychological demands placed on the learner may be different, but it is necessary to understand how they affect processing and, ultimately, learning.

It is important to acknowledge that the feedback type used in this study, i.e., explicit corrections, might have contributed to the finding that there are no statistical differences between the two feedback timing conditions. Previous research (Yilmaz, 2012, 2013) has shown that immediate explicit corrections are more effective than other types of immediate CF (e.g., recasts), a difference that has been attributed to the explicit linguistic indicators included in explicit corrections that facilitate a shift of attention from meaning to form and a cognitive comparison between the error and correct alternative (Yilmaz, 2012; Yilmaz & Granena, 2015). It is possible that the provision of delayed feedback in the form of explicit corrections might have increased the effectiveness of delayed feedback to a level that was not

possible with other types of feedback investigated in previous research (Arroyo & Yilmaz, 2018; Henderson, 2020). Because explicit corrections include both the error and its reformulation and are relatively direct about the ungrammaticality of the learner's utterance, they might have promoted the mental comparison between the error and correct alternative more reliably. Future research is needed to clarify the role of explicit corrections in increasing the effectiveness of delayed feedback.

CONCLUSION

The present study aimed to fill a gap in research on feedback timing in SLA by exploring the relative effects of immediate and delayed feedback on the acquisition of English –ing/-ed participial adjectives in video-based CMC. Methodologically, this study aimed at controlling for the contextualization and spacing of feedback instances that characterize immediate-feedback conditions and that previous research had failed to keep constant in delayed-feedback conditions. The results showed no effects for feedback or feedback timing on a GJT, but comparable effectiveness of immediate and delayed feedback on an OPT. These results indicate that the facilitative effect of delayed feedback on L2 learning may be greater than previously reported in the SLA literature. This has important pedagogical implications for distance language teaching contexts where feedback cannot be provided in real time to learners. In these contexts, the provision of video-based delayed feedback in the way that was provided in the present study may be more effective than the provision of delayed feedback in isolated instances. Additionally, other types of instructional settings such as face-to-face or hybrid language teaching could also benefit from adopting this way of providing delayed feedback on the recordings of learner-to-learner interactive oral tasks where feedback could be provided to the learners without needing to interrupt their oral interactions. However, there are some limitations to the generalizability of these results. This study examined a single language structure and a single feedback type in a single modality. Also, and despite the

strengths of experimental studies, sample size was small and, therefore, a larger sample size may have been needed to detect greater differences.. Finally, the design did not include a delayed posttest and, as a result, we could not determine the extent of knowledge retention in the long term.

Further research is needed that investigates different modalities in feedback timing conditions (e.g., text-based vs. video-based), as well as different feedback types and potential interactions between these factors. Practically, this research can inform about those conditions under which delayed feedback can be an effective alternative to immediate feedback and, theoretically, it can help shed more light on the cognitive mechanisms involved in the processing of different types and timing of feedback.

ACKNOWLEDGMENTS

NOTE

1. As an anonymous reviewer suggested, the lack of an effect on the GJT could be due to the difficulty of the test. The GJT, which included an error correction component in this study, might have been a more difficult test than the OPT, which provided participants with a concept word to guide their production.

REFERENCES

Arroyo, D. C., & Yilmaz, Y. (2017). The role of language analytic ability in the effectiveness of different feedback timing conditions. *Expanding individual difference research in the interaction approach: Investigating learners, instructors, and other interlocutors* (pp. 71–97). John Benjamins.

Arroyo, D. C., & Yilmaz, Y. (2018). An open for replication study: The role of feedback timing in synchronous computer‐mediated communication. *Language Learning, 68*(4), 942–972.

Baralt, M. (2013). The impact of cognitive complexity on feedback efficacy during online versus face-to-face interactive tasks. *Studies in Second Language Acquisition*, *35*(4), 689–725.

Canals, L., Granena, G., Yilmaz, Y., & Malicka, A. (2020). Second language learners' and teachers' perceptions of delayed immediate corrective feedback in an asynchronous online setting: An exploratory study. *TESL Canada Journal*, *37*(2), 181–209.

DeKeyser, R. (2007). Skill acquisition theory. In B. VanPatten & J. Williams (Eds.), *Theories in second language acquisition* (pp. 97–113). Lawrence Erlbaum.

Doughty, C. J., & Long, M. H. (2003). Optimal psycholinguistic environments for distance foreign language learning. *Language, Learning & Technology*, *7*(3), 50–80.

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

Ellis, R. (1991). *Instructed second language acquisition: Learning in the classroom.* Wiley-Blackwell.

Ellis, R. (2005). Measuring implicit and explicit knowledge of a second language: A psychometric study. *Studies in Second Language Acquisition, 27*, 141–172.

Gass, S. M., & Mackey, A. (2006). Input, interaction and output: An overview. *AILA Review, 19,* 3–17.

George, D., & Mallery, P. (2010). *SPSS for Windows step by step: A simple guide and reference 17.0 Update.* 10th Edition, Pearson.

González-Lloret, M. (2014). The need for needs analysis in technology-mediated TBLT. In M. Gonzalez-Lloret & L. Ortega (Eds.), *Technology and tasks: Exploring technology-mediated TBLT* (pp. 23–50). John Benjamins.

Goo, J., & Mackey, A. (2013). The case against the case against recasts. Studies in Second Language Acquisition, 35(1), 127–165.

Goo, J., Granena, G., Yilmaz, Y., & Novella, M. (2015). Implicit and explicit instruction in L2 learning. In P. Rebuschat (Ed.), *Implicit and explicit learning of languages* (pp. 443–482). John Benjamins.

Henderson, C. (2020). *Perfect timing? Exploring the effects of immediate and delayed corrective feedback, communication mode, and working memory on the acquisition of Spanish as a foreign language*. Unpublished PhD dissertation, Indiana University.

Henderson, C. (2021). The effect of feedback timing on L2 Spanish vocabulary acquisition in synchronous computer-mediated communication. *Language Teaching Research, 25*(2), 185–208.

Hrastinski, S. (2008). Asynchronous & synchronous E-learning. *Educause Quarterly*, *31*(4), 51–55.

Johnson, S., & Aragon, S. (2003). An instructional strategy framework for online learning environments. *New Directions for Adult and Continuing Education*, *2003*(100), 31–43.

Larsen-Freeman, D., & Celce-Murcia, M. (2015). The Grammar Book: Form, Meaning and Use for English Language Teachers (3rd edition). National Geographic

Learning/Cengage Publishing Company.

Li, S. (2010). The effectiveness of corrective feedback in SLA: A meta-analysis. *Language Learning, 60*(2), 309–365.

Li, S., Shintani, N., & Ellis, R. (Eds.). (2015). Synthesizing research in applied linguistics: the complementary contributions of narrative reviews and meta-analyses, *Special Issue of Applied Linguistics, 36*(3).

Li, S., Zhu, Y., & Ellis, R. (2016). The effects of the timing of corrective feedback on the acquisition of a new linguistic structure. *Modern Language Journal*, *100*(1), 276–295.

Loewen, S., & Erlam, R. (2006). Corrective feedback in the chatroom: An experimental study. *Computer Assisted Language Learning, 19*(1), 1–14.

Long, M. H. (1996). The role of the linguistic environment in second language acquisition. In W. R. Ritchie & T. J. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 413–468). San Diego: Academic Press.

Long, M. H. (2007). *Problems in SLA* . Lawrence Erlbaum.

Long, M. H. (2015). *Second language acquisition and task-based language teaching.* Wiley-Blackwell.

Lyster, R., & Ranta, L. (2013). Counterpoint piece: The case for variety in corrective feedback research. *Studies in Second Language Acquisition, 35*(1), 167–184.

Lyster, R., & Saito, K. (2010). Oral feedback in classroom SLA. *Studies in Second Language Acquisition*, *32*(2), 265–302.

Lyster, R., Saito, K., & Sato, M. (2013). Oral corrective feedback in second language classrooms. *Language Teaching*, *46*(1), 1–40.

Mackey, A., & Goo, J. (2007). Interaction research in SLA: A meta-analysis and research synthesis. In A. Mackey (Ed.), *Conversational interaction in second language acquisition: A series of empirical studies* (pp. 407–453). Oxford University Press.

Monteiro, K. (2014). An experimental study of corrective feedback during video-conferencing. *Language Learning & Technology*, *18*(3), 56–79.

Nassaji, H. (2015). Qualitative and descriptive research: Data type versus data analysis. *Language Teaching Research, 19*(2), 129–132.

Nicholas, H., Lightbown, P. M., & Spada, N. (2001). Recasts as feedback to language learners. *Language Learning*, *51*(4), 719–758.

Pica, T. (1988). Interlanguage adjustments as an outcome of NS-NNS negotiated interaction. *Language Learning, 38*(1), 45–73.

Plonsky, L. & Oswald, F. (2014). How big is "big"? Interpreting effect sizes in L2 research. *Language Learning*, *64*(4), 878–912.

Quinn, P. (2014). *Delayed versus immediate corrective feedback on orally produced passive errors in English* (Unpublished doctoral dissertation). University of Toronto, Toronto, Canada.

Rassaei, E. (2017). Video chat vs. face-to-face recasts, learners' interpretations and L2 development: A case of Persian EFL learners. *Computer Assisted Language Learning, 30*(1-2), 133-148. 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). John Benjamins.

Saito, K., & Akiyama, Y. (2017). Video‐based interaction, negotiation for comprehensibility, and second language speech learning: A longitudinal study. *Language Learning*, *67*(1), 43–74.

Sauro, S. (2009). Computer-mediated corrective feedback and the development of L2 grammar. *Language Learning & Technology*, *13*(1), 96–120.

Scovel, T. (1974). 'I am interesting in English'. English Language Teaching Journal, 28(4),

305–312.

Sheen, Y., & Ellis, R. (2011). Corrective feedback in language teaching. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 593–610). Routledge.

Shintani, N., & Aubrey, S. (2016). The effectiveness of synchronous and asynchronous written corrective feedback on grammatical accuracy in a computer‐mediated environment. *Modern Language Journal*, *100*(1), 296–319.

Smith, S. M., Handy, J. D., Angello, G., & Manzano, I. (2014). Effects of similarity on environmental context cueing. *Memory*, *22*(5), 493–508.

Tulving, E. (1972). Episodic and semantic memory. In E. Tulving & W. Donaldson (Eds.), *Organization of memory* (pp. 381–403). Academic Press.

Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, *80*(5), 352–373.

Wheeler, R. L., & Gabbert, F. (2017). Using self-generated cues to facilitate recall: A narrative review. *Frontiers in Psychology*, *8*, 18–30.

Yilmaz, Y., & Yuksel, D. (2011). Effects of communication mode and salience on recasts: A first exposure study. *Language Teaching Research*, *15*(4), 457–477.

Yilmaz, Y. (2012). The relative effects of explicit correction and recasts on two target structures via two communication modes. *Language Learning, 62*(4), 1134–1169.

APPENDICES

Appendix A. Control 1 and 2 Group Comparisons.

A mixed-design ANOVA with Group as the between-subjects factor and Time as the within-subjects factor was conducted for each outcome measure. In the case of the GJT (see Table 5), the pretest scores of the two control conditions were comparable, $F(1, 18) < .01$, $p =$

.990, $\eta^2 < .001$, and the mixed-design ANOVA further showed that none of the effects were statistically significant: Group, $F(1, 18) = .09$, $p = .776$, $\eta^2 < .01$, Time, $F(1, 18) = 4.19$, $p = .055$, $\eta^2 = .19$, and interaction between Time and Group, $F(1, 18) = .29$, $p = .596$, $\eta^2 = .01$.

Table 5

Descriptive Statistics for the Grammaticality Judgment Test

Group		Pretest		Posttest		Gains	
	N	*M*	*SD*	*M*	*SD*	*M*	*SD*
Control 1	10	0.21	0.25	0.37	0.34	0.16	0.36
Control 2	10	0.21	0.25	0.3	0.31	0.09	0.16

Similarly, for the OPT (see Table 6), there were no statistically significant differences in the pretest scores of the two control conditions, $F(1, 18) = 3.30$, $p = .086$, $\eta^2 = .155$, and the mixed-design ANOVA further showed that none of the effects were statistically significant: Group, $F(1, 18) = 2.96$, $p = .103$, $\eta^2 = .14$, Time, $F(1, 18) = .39$, $p = .050$, $\eta^2 = .19$, or the interaction between Time and Group, $F(1, 18) = .48$, $p = .494$ $\eta^2 = .02$.

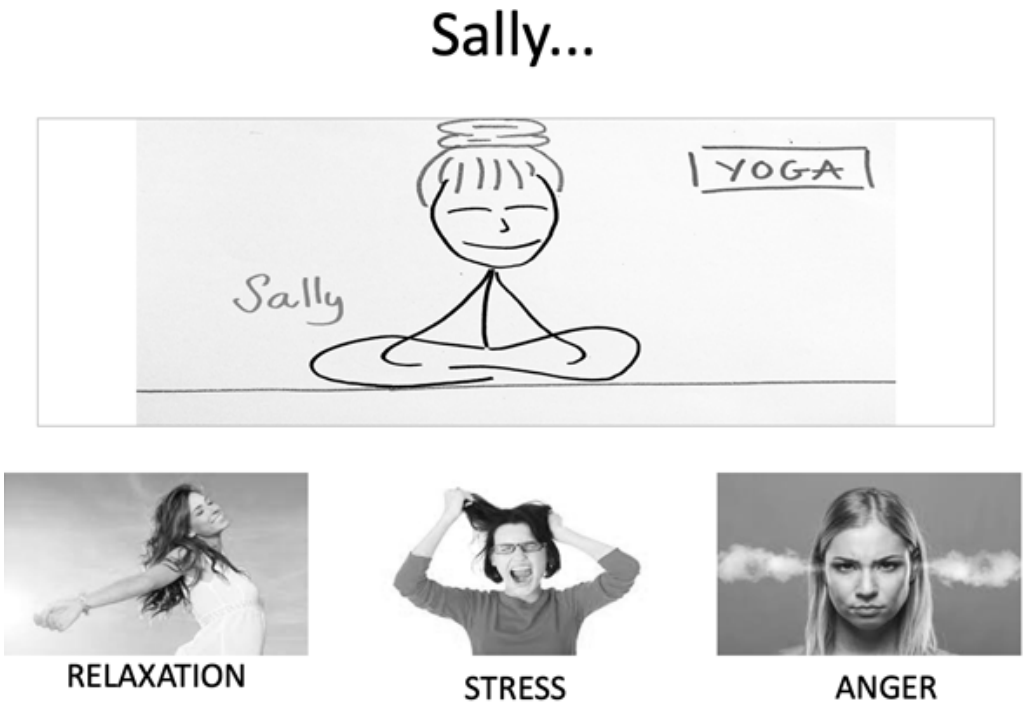
Table 6

Descriptive Statistics for the Oral Production Task

Group		Pretest		Posttest		Gains	
	N	*M*	*SD*	*M*	*SD*	*M*	*SD*
Control 1	10	0.46	0.25	0.51	0.24	0.05	0.19
Control 2	10	0.29	0.15	0.39	0.14	0.1	0.13

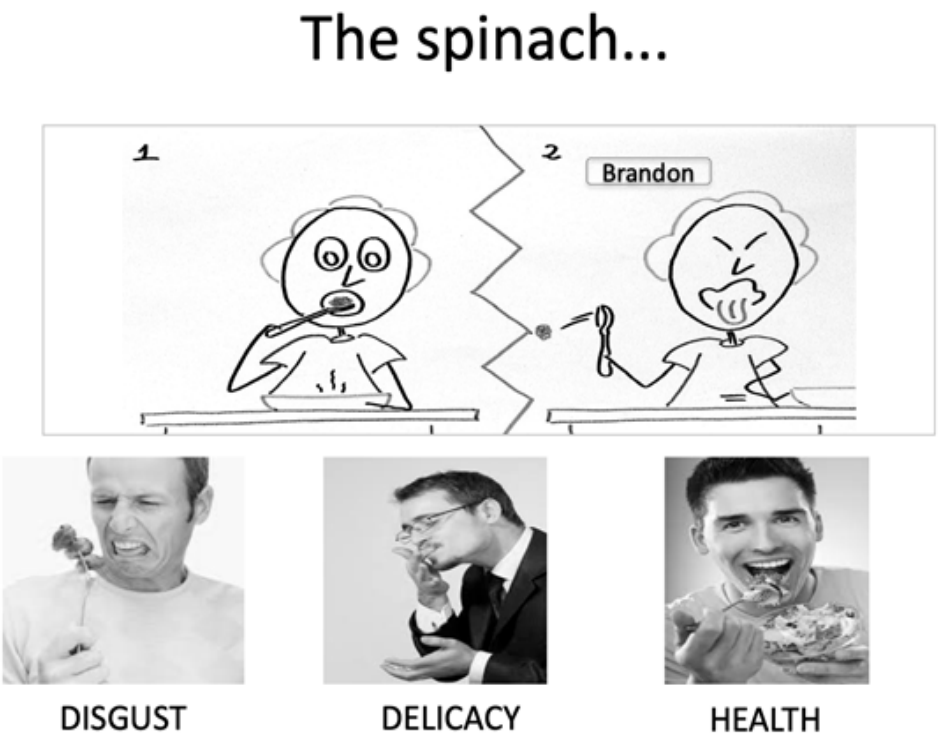
Appendix B. Sample Item from the Treatment Task for –ed Adjective (Experimenter's Slide).

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

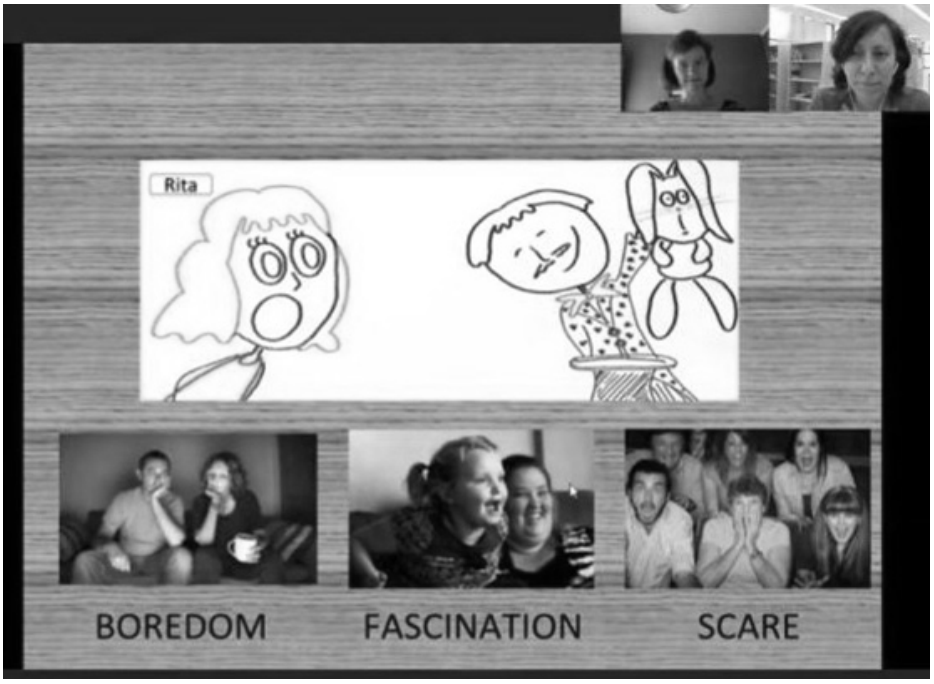


Appendix D. Sample Item from the Treatment Task for –ing Adjective (Experimenter's Slide).

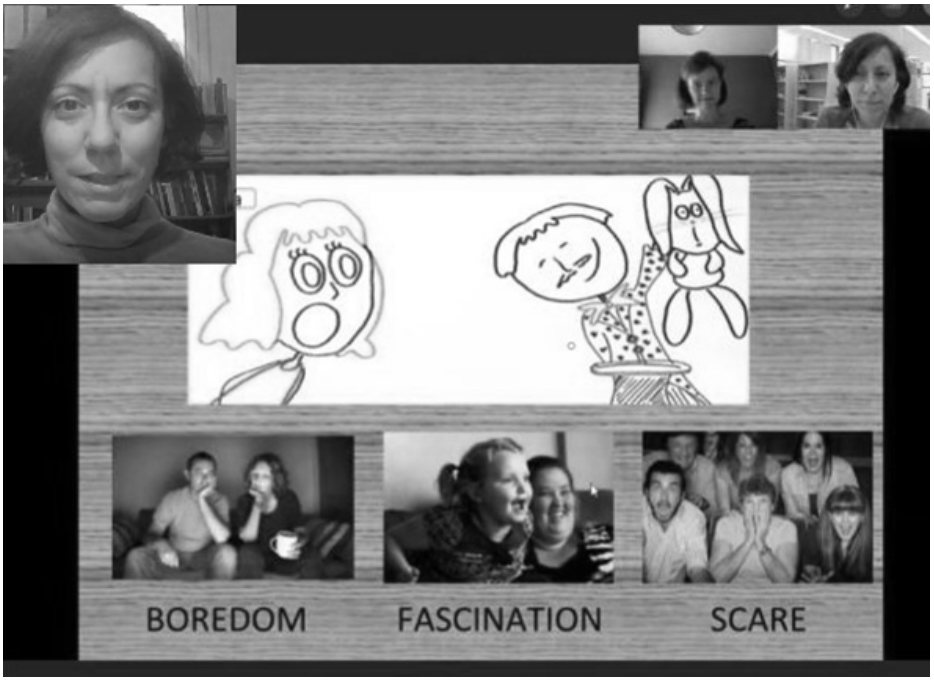
Appendix E. Sample Item from the Treatment Task for –ing Adjective (Learner's Slide).



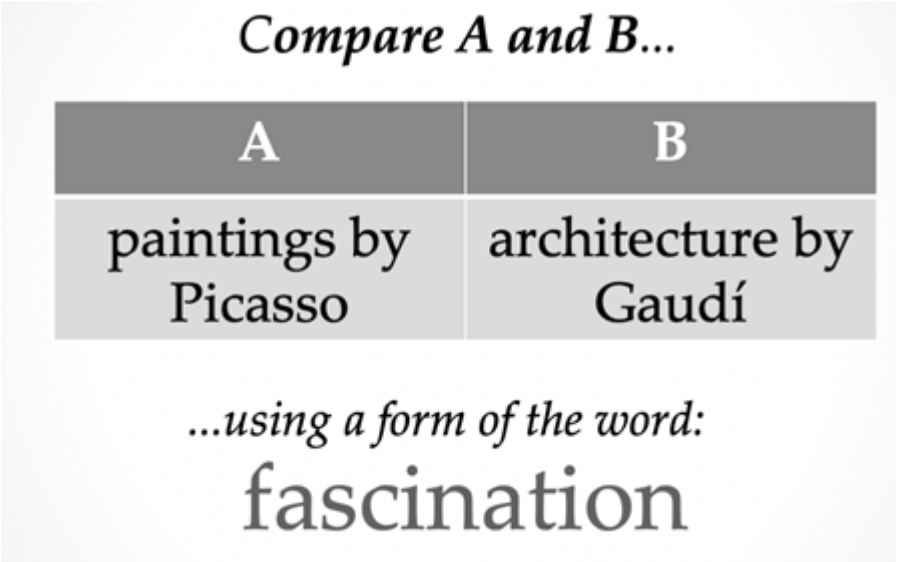
Appendix F. Screen Capture of the Immediate-feedback Condition.



Appendix G. Screen Capture of the Delayed-feedback Condition.



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



Appendix I. Descriptive Statistics for Grammatical Items on the Grammaticality Judgment Test.

Table 7

Descriptive Statistics for Grammatical Items on the Grammaticality Judgment Test

Group		Pretest		Posttest		Gains	
	N	*M*	*SD*	*M*	*SD*	*M*	*SD*
Control	20	0.98	0.04	0.93	0.10	-0.05	0.12
Immediate Feedback	16	0.91	0.15	0.92	0.14	0.01	0.14
Delayed Feedback	16	0.98	0.05	0.98	0.07	0.00	0.09