

The Role of Corrective Feedback Timing in Task Engagement and Oral Performance

Javad Belali¹ and Karim Sadeghi^{2*}

¹ *Valfajr Street Faculty of Humanities*

2Professor, English Department School of Humanities, Urmia University, Urmia, Iran

Received: 2019/03/14

Accepted: 2019/05/11

Abstract: Corrective Feedback (CF) provision on performance involves calling attention to learners' erroneous utterances, which stimulates classroom language learning. Despite widespread research in this area, a controversy still exists as to the time of CF in communicative tasks especially when learners' task engagement is targeted. Employing 60 intermediate-level Iranian learners, a four-week study examined the effects of immediate versus delayed CF types on oral engagement, grammar gain as well as on Complexity, Accuracy and Fluency (CAF) in the context of a dictogloss task. The statistical analysis showed that despite statistically significant improvements in the case of Grammaticality Judgement Test (GJT) from the pre-test to the post-test, no statistically significant difference was found between the two groups with regard to their grammar gain, complexity, accuracy, fluency, as well as on their social and cognitive engagement. However, results in the case of behavioral and emotional engagement revealed that there were statistically significant differences between the two groups of immediate and delayed CFs. The findings of the present study call for more flexibility and confidence on the part of the teachers in providing either immediate or delayed CF whenever appropriate and without being concerned about the adverse effects of feedback.

Keywords: Accuracy, Complexity, Corrective Feedback, Delayed Feedback, Fluency, Immediate Feedback, Oral Engagement.

* Corresponding Author.

Authors' Email Address:

¹ Karim Sadeghi (k.sadeghi@urmia.ac.ir), ² Javad Belali (belali.javad@yahoo.com),

ISSN (Online): 2322-5343, ISSN (Print): 2252-0198 © 2020 University of Isfahan. All rights reserved

Background

Introduction

The idea of what it means to be proficient in a language has received due research attention in recent years (Housen, Kuiken, & Vedder, 2012). From the diverse pack of research, complexity, accuracy and fluency (CAF) emerged as distinct elements of second language (L2) performance (Ellis & Barkhuizen 2005; Housen & Kuiken, 2009). Identifying factors that shape the development of CAF in L2 learning is a concern for both researchers and teachers. Many researchers believe that the development of CAF can be related to a number of factors such as motivation and corrective feedback (Ellis, 1994; Housen et al, 2012).

Corrective feedback (CF) is a frequent practice in the field of education and learning in general and second language education in particular. Research on oral corrective feedback has received serious attention over the last 20 years. Many investigations have been conducted to show the benefits of CF and learner uptake in L2 development in classroom contexts (e.g.; Lee 2013; Li, 2010; Lyster & Saito, 2010; Lyster, Saito & Sato, 2013). However, little attention has been paid to how learners respond to the timing of CF.

The ideas of the timing of CF have been greatly influenced by reform movements in thinking about new approaches to language learning and teaching. The answer to the question 'How should corrective feedback be provided?' is a radical departure from the belief that corrective feedback should always be provided immediately (Hendrickson, 1978). From the Interactionists' eyes, errors are the natural by-product of development in the process of L2 acquisition, and that the optimal time for the treatment of errors is when those utterances have not secured the hearer's attention in communication.

Two concepts of social justice and responsibility of the instructors are of utmost importance in providing feedback on learners' performance in oral tasks. For example, regarding the social justice, the trite provision of CF by teachers creates negative evidence in learners' mind such that they have never belonged to the class (Hattie & Timperley, 2007). Salend (2001) also claimed that if teachers make comparisons when providing feedback, they should make comparison between individual learner's current and prior performance rather than compare it with his/her peers' performance. This establishes real standards, eliminating comparisons that learners might find disappointing and discouraging. In their social responsibilities, teachers should be brought to a clear understanding in their duty of how to correct learners' errors. Therefore, learners' inner and outer changes are based on the labors of their teachers. Additionally, teachers usually require student participation and oblige

students to speak in classroom discussions. Such requirements include having students who are engaged and motivated to learn (Rocca, 2008).

When language teachers are implicit about learners' inaccurate language use, learners typically infer that teachers' silence indicates that they have made no errors. If teachers correct their errors immediately, this correction discourages them from speaking. This is particularly the turning point that the issue of engagement raises.

In line with Stroud's (2015) definition, the term 'engagement', refers to “devoting a lot of effort, remaining highly focused, persisting and enjoying challenges” (p. 2). The true value of a student engagement is determined primarily by the measure and the sense in which s/he has attained liberation from the self. The term 'engagement', in meaning-focused interaction, is to see with one's own eyes and to judge one's performance in tasks without succumbing to the suggestive power of the peers or fellow students. This complex phenomenon can be classified into four types: Behavioral (the amount of the delivery of messages and actual performance), Cognitive (mental efforts and the level of thoughtfulness and 'awareness'), Social (the degree of learners' extrovertsiveness to their interlocutors during performance), and Emotional (having positive or negative 'attitude' towards tasks and learners) (Philp & Duchesne, 2016).

It is not easy to comprehend why we feel constrained to call the propositions of engagement 'true'. Engagement ideas correspond to more or less exact issues such as confidence, interest, and willingness to communicate, the last of which is assumed to be the exclusive cause of the genesis of those ideas. Likewise, the practice of seeing conditions such as low levels of confidence and interest in tasks can lead to low levels of engagement (MacIntyre, Dörnyei, Clément & Noels, 1998). As a result, diagnosing these conditions is of utmost importance for ensuring that learners are progressing.

In addition, an adverse effect of the almost sole focus on CF type is that, in contemporary CF research, it seems a priori that CF is to be provided immediately. However, the positive outcomes from the very few investigations of the empirical effects of delayed CF (Hunter, 2011; Sheen, 2012) propose some opposing evidence to that belief. Therefore, more research is required to investigate whether there is any difference in the L2 development that results from immediate and delayed CF. Furthermore, if second language acquisition (SLA) research is to help not only learners and teachers but also researchers, then certainly researchers should study whether there are any empirical facts that can be obtained to support the widely unsupported recommendations about the timing of CF in the instructional literature. This research is

accordingly an endeavor to investigate whether the timing of CF would have any effect on grammar gain, task engagement, and CAF measures of oral production.

Literature Review

From a historical perspective, the issue of the timing of corrective feedback on L2 speaking development has long been reduced to an absolute minimum and attention to it has only recently re-emerged (Quinn, 2014). In fact, contemporary researchers have collected a large quantity of empirical data indicating that the timing of CF facilitates the progression of L2 knowledge.

Rahimi and Dastjerdi (2012), for example, documented a comparative study of the L2 development, focusing on immediate and delayed CF, on Iranian EFL learners' complexity, accuracy and fluency in oral tasks. Learners, who benefited from both immediate and delayed CF, were divided into two groups of ten. Results showed that delayed feedback has a statistically positive effect on accuracy and fluency, but not on complexity. Unfortunately, the low number of the participants in this study did not guarantee the validity of the results.

Along similar lines, Varnosfadrani (2006) looked into the development effects of delayed versus immediate CF, in the form of both explicit versus implicit CF, on 28 adult Iranian EFL learners. Participants took part in two text reconstruction tasks. In both tasks participants were provided with explicit correction. Tailor-made tests (tests containing items showing errors made by learners during an activity) (LaPierre, 1995) were made for each error. Statistically significant improvement was found on the errors that had been made, but there were no statistically significant differences between the improvements that resulted from immediate or delayed CF. However, the outcomes of Varnosfadrani's study cannot be attributed to a manipulation in the timing of CF because it appears that the CF types have been confounded with CF timing which prevents a true comparison of the effects of timing on CF.

Another well designed study was done by Quinn (2014). In this study, 90 intermediate-level adult ESL learners were randomly assigned to immediate, delayed, and no CF conditions. The target grammatical structure was English passive. In a pre-test–immediate post-test–delayed post-test design including an aural grammaticality judgment test, a written error correction test and an oral written error correction test, the learners took part in three communicative tasks. The immediate and delayed feedback consisted of a prompt that forced the learners to self-correct followed by a recast. Results revealed that there were statistically significant improvements on all three conditions, but there were no statistically significant

differences among groups between conditions. However, a main problem of this study is that it was laboratory-based. Therefore, the potential generalizability of the findings to classroom contexts is limited.

Li, Zhu and Ellis (2016), reported a comparative study of the effectiveness of the immediate versus delayed CF on the acquisition of past passive structure. Employing 120 learners from 4 intact classes, learners were assigned to immediate, delayed, task-only, and control groups. The study used a pre- to post-test design by employing a Grammaticality judgement test and an elicited imitation test as pre/post-test, two dictogloss tasks and recasts as a CF provision type for the two immediate and delayed groups. Results implied that the participants benefitted more from the immediate CF than delayed provision. However, the problem here lies in the likelihood of a practice effect in the data collected from the control group who received different instruction and took part in the same tests employed in the study.

Additionally, as learner engagement is an essential link that combines the provision of CF with learning outcomes, research on this area can contribute to the current theorizing about the oral CF mechanism. According to Ellis (2010), learner engagement is associated with the ways that learners respond to CF.

Research on learners' perspectives on feedback only began to develop in the 1990s (e.g. Ferris, 1995; Hedgcock & Lefkowitz, 1996). The focus of the research has mainly been on learners' preferences and attitudes based on questionnaire surveys (e.g., Cohen, 1987; Ferris, 1995), and little or no attention has been paid to actual student reactions to teacher feedback in classroom contexts. Adopting a more inclusive view of how feedback affects engagement with tasks, Butler and Winne (1995) suggested that to better direct learning in authentic tasks, feedback should provide information about cognitive activities for learning. To their belief, monitoring is at the heart of self-regulated task engagement and the internal feedback it generates is critical in shaping the evolving pattern of a learner's engagement in a task. Nevertheless, apart from the theoretical nature, the focus of the study (Butler & Winne, 1995) was merely on cognitive engagement and failed to take into account other aspects of engagement (emotional, social and behavioral).

Research studies that focus on error correction time are extremely rare in comparison to the large number of studies that focus on error correction (Farmani, Akbari & Ghanizadeh, 2017). Also, despite the importance of psychological variables, there are no studies that delve into the effect of error correction on engagement. Although empirical attention to the timing

of CF on psychological variables has been limited in the contemporary SLA literature, the following findings addressed a gap in the timing of CF and engagement.

Farmani et al (2017) studied the L2 development that resulted from immediate and delayed CF on 90 intermediate Iranian EFL learners' motivation. The participants were randomly divided into three immediate, delayed and control groups. In a pre-test to post-test design, the three groups were given a motivation questionnaire and a Cambridge Proficiency Test in the first and the last session of the term to see the effect of different kinds of error correction (immediate and delayed) on learners' motivation. The results of the study showed that the immediate CF treatment was found to have led to a significantly better performance than the delayed CF treatment.

Another study was conducted by Shabani and Safari (2016) who reported the effects produced by immediate versus delayed feedback on learners' oral accuracy and anxiety. 100 Iranian EFL students aged from 13 to 30 took part in the study. The results revealed that although all the participants expressed some levels of anxiety when receiving error corrections, the level of their anxiety was significantly different and the immediate CF group experienced more anxiety than the delayed CF group. Moreover, results revealed that the level of anxiety facilitated the improvement of accuracy during speech.

In another study, Stroud (2017) conducted a well-designed investigation in an attempt to uncover the effect of task scoring and tracking on EFL university learners' engagement. Using card games as a measurement of the aspects of engagement, the results indicated that the cards had a significant effect on the behavioral engagement but had no effect on the social and cognitive engagement. Likewise, results obtained from High School Survey of Student Engagement did not identify any significant effects on learners' affective dispositions toward tasks. However, the major limitation of the study, according to Stroud (2017), is that the use of the card game may well have been motivating enough for the learners to engage more in discussions.

In summary, as the research on the importance of CF in L2 learning is growing rapidly, research in the case of oral engagement is still in its infancy. The current study has been intended to add to this research area and attempts have been made to prevent the weaknesses of the previous investigations by including the effects of corrective feedback on EFL learners' grammatical knowledge gain, oral engagement and on their complexity, accuracy and fluency in speaking.

In order to delve more deeply into the aforementioned issues, the following questions are proposed:

1. Is there any significant difference between delayed and immediate CF in their effect on the grammatical knowledge gains of learners?
2. Do delayed and immediate CF engage learners differently in performing an oral task?
3. Is there any significant difference between delayed and immediate CF in their effect on the complexity of learners' oral production?
4. Is there any significant difference between delayed and immediate CF in their effect on the accuracy of learners' oral production?
5. Is there any significant difference between delayed and immediate CF in their effect on the fluency of learners' oral production?

Methodology

Design of the Study

This quasi-experimental study followed the pre-test/post-test design to investigate the effects of immediate and delayed CF on improvements in learners' English past passive structure in terms of CAF and task engagement. The treatments were proceeded with the random division of participants into immediate and delayed conditions. Each participant followed a 4-week treatment process, as outlined in Table 1. Finally, quantitative procedures were employed to analyze the elicited data.

Participants

A convenience sample of 60 intermediate level learners (aged from 12 to 18) was recruited from Jihad private EFL institutes in Urmia, Iran. The intermediate level learners were selected because they were believed to present the greatest potential to exhibit development in the target form from CF treatment. The participants were randomly divided into two groups of immediate and delayed CF (N = 30- in each group). This consisted of 27 female students (13 within immediate group and 14 within delayed group) and 33 male students (17 within immediate group and 16 within delayed group). The students met twice a week for 90 minutes during which they were engaged in oral tasks. Two teachers also participated in the study, having 7 and 10 years of teaching experience.

To ensure that groups matched each other in terms of their initial competence of the target structure, both groups took a Grammaticality Judgement Test (GJT) at the outset of the

study. Further, the homogeneity of the participants in this study was established from the perspectives of cultural, ethnic and language background, with Azeri being their mother tongue and Persian as their second language. Finally, since previous research has shown that corrective feedback leads to better performance compared to no feedback (e.g. Li, Zhu, & Ellis, 2016; Loewen & Nabei, 2007; Quinn, 2014), it was decided not to use a control group.

Instruments

The instruments used for gathering the data were a researcher constructed Grammaticality Judgement Test (GJT), the High School Survey of Student Engagement (HSSSE) and a dictogloss task. The GJT was designed with the aim of determining participants' current level of competence in the target structure and was divided into three parts. Using a Nominal Scale to judge the grammaticality of the sentences (correct or incorrect), we provided the participants with 30 sentences, including 20 target passive items (10 correct and 10 incorrect) and 10 distractor items with an active verb form. Participants' answers to the GJT (used both as pre-test and post-test) were scored by assigning 'one' point for correct responses and 'zero' for all other responses, including 'not sure'. The reliability of GJT was assessed by KR-20 index, which indicated an acceptable level of reliability ($r = 0.96$).

A dictogloss task was designed to stimulate the use of the 'past passive' structure that was the focus of the current study. The past passive structure was selected because it is a structure acknowledged to require a long time for L2 learners to acquire (Larsen-Freeman, 1995). The dictogloss task comprised 3-5-minute pair discussions, with one-to-one performance of the learners to the teacher and the whole class to evaluate their ability to state opinions during their task performance. Pair work discussions were not directly observed by the teacher. All performances were recorded with the help of mobile phones placed on students' desks.

Accordingly, in order to make sure that the timing for the two conditions would be the same, it was necessary to pilot the instructional task under the two conditions. 10 participants were called up for recording the pilot sessions with 5 participants in each condition. After reviewing the recorded speech from the pilot sessions, only small changes were made (the length of the text for the dictogloss task was reduced and adapted).

High School Survey of Student Engagement (HSSSE) (Yazzie-Mintz, 2007). is a well-established engagement survey and a widespread examination of student engagement and school environment problems designed to measure cognitive, social and emotional engagement

of students in learning during and outside of the class. The survey consists of 31 items with questions like 'Have you ever considered dropping from this school?', 'Have you ever got bored in class?' with the choices ranging from 'Strongly Disagree' to 'Strongly Agree'.

The adjusted version of HSSSE that was used aimed at measuring self-reported emotional engagement of students in learning in class time. The 13-item survey focused on learners' affection in terms of (1) the way students feel about their level in class, and the teachers within their classroom; and (2) affections toward association or disassociation of students with their peers in the class and, mainly, their task interlocutors. The reliability of HSSSE was assessed by KR-20 index, which indicated an acceptable level of reliability ($r = 0.50$).

Procedure

Following the curriculum of the institute, based on *Topnotch 2* book (Saslow, Ascher & Morsberger, 2011), the focus of grammatical structure was the past passive. The data collection took place over 4 weeks. In the first week, each participant took part in the GJT which aimed at measuring the knowledge of English past passive structure. Participants were provided with an example (but not an example from the target structure) to make sure that they understood the purpose of the test.

Next, the 15-minute instruction for both immediate and delayed CF group in the second week was carried out in the following way: The participants were shown a picture of 'a window' and 'a boy having a stone in his hand'. They then listened to the instructor while they read the same sentence written below the picture: *the boy broke the window*. Later, the subject, object, verb and the receiver of the action were identified and explained to the students. The instructor then clarified that the verb was in the active form and explained that the subject in the active sentences is responsible for the action of the verb. Afterward, the participants were presented with the same picture (the broken window) but with the picture of the agent (the boy) missing. The following sentence under that picture was then read by the instructor: *the window was broken by the boy*. The instructor asked the participants to tell the difference between the two sentences, clarifying that just the structure of the sentence had changed, not the meaning.

When the instruction finished, the participants were asked to take part in a 20-25 minute dictogloss task with the purpose of eliciting the use of past passive structure. The participants in this task were at first presented with personal questions relating to the topic of the task such as

'How often do you paint your house inside?' in order to attract their interest. They then listened to a narrative told by the instructor. The flow of information was controlled by allowing learners to discuss in pairs (about 3-5 minutes) so that the focus of the task was meaningful while the context was created for the participant's use of the past passive structure. After the pair work, they were invited to describe the narrative to the group in cooperation with their partners. Retelling the narrative gave the participants an opportunity to make errors in the formation of passive structure on which either immediate or delayed CF was provided. The two groups were provided with an explicit correction (i.e. when the teacher provides the correct form, s/he clearly indicates that what the student had said was incorrect). Since previous research supports the priority of the explicit correction in the process of L2 learning (e.g. Lee, 2013) over recasts (reformulating learners' incorrect utterances while maintaining a focus on meaning), explicit correction was used for correcting learners' erroneous utterances in both groups. Moreover, in the immediate CF provision, the correction was provided at the moment that an error occurred, while in the delayed CF provision, correction was postponed until the task was done and after the participants' suggestions for correcting each other's errors. Also, CF was given merely on errors produced on the passive structure.

On week 4, the participants in both immediate and delayed CF group were examined again with the same GJT that they had taken in the first week. Finally, HSSSE was given in week 3 for immediate and delayed CF groups in order to measure their affective dispositions. All performances were recorded with the help of mobile phones placed on each pairs of students' desks.

Table 1 represents the summary of the research process.

Table 1. Overview of Research Schedule

<i>Session 1</i>	<i>Session 2</i>	<i>Session 3</i>	<i>Session 4</i>
Pre-test GJT (15-20 minutes)	15-minute instruction for English passive lesson 20-25 Dictogloss task (immediate/delayed CF groups)	HSSSE (immediate/ delayed CF)	Post- test GJT (15-20 minutes)

Data Analysis

To examine the participants' CAF, the researcher analyzed the audio-recorded classes by counting 'target like use of verb tenses', 'error free clauses' and 'number of self-corrections' for measuring Accuracy, 'number of pauses' and 'number of repetitions' for Fluency, and

finally, by counting 'frequency of use of prepositions' and 'frequency of use of conjunctions' for Complexity (Ellis, 2003; p. 117).

Behavioral engagement in audio-recorded dictogloss task was analyzed by counting 'total words spoken'. Social and cognitive engagement were measured by means of 'stated opinions', 'agreements', 'disagreements', 'paraphrasing' and 'help given'. Finally, analysis from HSSSE was used as a measurement scale for emotional engagement.

Findings

Overall. Prior to conducting the statistical analyses, assumptions for normality and homogeneity were checked. Normality was violated by all conditions on all tests, with the exception of the behavioral and emotional engagement. Therefore, the current research proceeded with both parametric and non-parametric analyses.

GJT Results

To answer the first research question, it was necessary to determine whether there was a change in grammatical knowledge of the learners in their performance from pre-test to post-test. Therefore, Wilcoxon Signed Ranks Test was employed for analyzing each CF group's gain from the pre-test to the post-test stage and a Man Whitney U test was used for comparing the post-tests obtained for comparing differences in performance between immediate and delayed CF groups.

Immediate CF Pre/Post-test Results

The pre-test and post-test in the immediate CF group was examined for the aim of supporting the improved performance in the target structure. The results obtained from Wilcoxon Signed Ranks Test indicated a statistically significant difference between the pre-test to the post-test of the immediate CF group.

Table 2. *Pre/post-test Results of Immediate CF*

<i>scores</i>					
		<i>N</i>	<i>Mean Rank</i>	<i>Z</i>	<i>Asymp. Sig. (2-tailed)</i>
Post-test – Pre-test	Negative Ranks	2 ^a	1.75	-4.571 ^d	.000 ^e
	Positive Ranks	26 ^b	15.48		
	Ties	2 ^c			
	Total	30			

- a. Post-test < Pre-test
- b. Post-test > Pre-test
- c. Post-test = Pre-test
- d. Based on negative ranks
- e. Wilcoxon Signed Ranks Test; $P < 0.01$

Delayed CF Pre/post-test Results

The Wilcoxon Signed Ranks Test showed a statistically significant effect for time ($p = 0.00$). The results obtained from Wilcoxon Signed Ranks Test indicated that the scores of the learners were different across the two time points, that in general there were statistically significant differences between the delayed CF groups' pre-test and post-test scores. The results are presented in Table 3, which show that the delayed feedback group performed better on the post-test compared to the pre-test.

Table 3. *Pre/post-test Results of Delayed CF Group*

<i>Scores</i>		<i>N</i>	<i>Mean Rank</i>	<i>Z</i>	<i>Asymp. Sig. (2-tailed)</i>
Post-test – Pre-test	Negative Ranks	0 ^a	.00	-4.789 ^d	.000 ^e
	Positive Ranks	30 ^b	15.50		
	Ties	0 ^c			
	Total	30			

- a. Post-test < Pre-test
- b. Post-test > Pre-test
- c. Post-test = Pre-test
- d. Based on negative ranks.
- e. Wilcoxon Signed Ranks Test; $P < 0.01$

Immediate/Delayed Post-test Comparisons on the Grammaticality Judgment Test

The descriptive scores presented in Table 4 to establish the comparability between the immediate and delayed CF groups. The results of the pre-test pointed us towards the fact that learners in both groups possessed almost no knowledge of the past passive structure ($M=11.20$, immediate CF; $M = 9.17$, delayed CF out of 20).

Despite statistically significant increase in scores between groups on their post-test scores, results obtained from the Mann-Whitney U test revealed no statistically significant difference between the two groups, although in post-test, delayed CF group scored much higher ($M = 34.32$) than immediate group ($M = 26.68$). Therefore, there was no statistically significant difference between the two treatment groups in terms of their effect on grammatical knowledge gain.

Table 4. *Immediate/Delayed Post-test Comparison*

		Scores				
	Group	N	Mean Rank	Mann-Whitney U	Z	Asymp. Sig. (2-tailed)
Post-test	immediate CF	30	26.68	335.500	-1.748	.080
	Delayed CF	30	34.32			
	Total	60				

Note: $P < 0.05$

Engagement

Changes in Behavioral Engagement

Results obtained from the Kolmogorov-Smirnov test, revealed that the data collected from participants' behavioral engagement were normally distributed. The results reported in Table 5 show that the Sig value of the test is more than 0.05 ($p = 0.2$). Therefore, there were no meaningful differences between the distribution of the data and the normal distribution. In other words, the data was normally distributed.

Table 5 depicts the results:

Table 5. Normality of the Behavioral Engagement

One-Sample Kolmogorov-Smirnov Test			behavioral
N			60
	Normal Parameters ^{a,b}	Mean	17.72
		Std. Deviation	8.501
	Test Statistic		.085
	Asymp. Sig. (2-tailed)		.200 ^{c,d}

a. Test distribution is Normal.
 b. Calculated from data.
 c. Lilliefors Significance Correction.
 d. This is a lower bound of the true significance.

One condition of the independent sample *t*-test is the notion of 'the equality of variance' for the two groups, which is examined by the Fisher test (F). Therefore, the *p* value obtained from the F test is more than 0.05 ($p = 0.604$). Thus, no statistically significant difference was found between the two CF groups with regard to the equality of the variances. On the other hand, because the *p* value of the *t*-test is less than 0.01 (0.000), there existed a statistically significant difference between the immediate and delayed CF types.

Table 6. Comparison Analysis Results for Behavioral Engagement

		Group Statistics			
	Group	N	Mean	Std. Deviation	Std. Error Mean
Immediate vs Delayed	Immediate CF	30	23.43	6.735	1.230
	Delayed CF	30	12.00	5.837	1.066

Independent Samples Test

	Levene's Test for Equality of Variances				t-test for Equality of Means		
	<i>F</i>	<i>Sig.</i>	<i>t</i>	<i>df</i>	<i>Sig. (2-tailed)</i>	<i>Mean Difference</i>	<i>Std. Error Difference</i>
Equal variances assumed	.272	.604	7.027	58	.000	11.433	1.627

Note: $P < 0.01$

Changes in Social and Cognitive Engagement

Within the two weeks of experiment, performance data was analyzed in order to measure whether there existed changes in social and cognitive engagement. Table 7 reports the descriptive statistics: a) Delayed CF group in the case of social and cognitive engagement ($M = 33.50$) with regard to the 'opinions, agreements, disagreements' outperformed immediate CF group ($M = 27.50$), b) Delayed CF ($M = 32.00$) group in the case of 'Question, clarify, request' was better than immediate CF group ($M = 29$). Although the delayed CF group showed higher scores than the immediate CF group, the differences did not reach statistical significance ($p = 0.117$; $p = 0.440$).

Besides, results indicated that there existed no statistically significance difference between the two groups ($p = 0.154$) in the case of 'paraphrase and help given' although immediate CF group did better ($M = 33.17$) than the delayed CF group ($M = 27.83$).

Table 7. Comparison analysis results for social and cognitive engagement

<i>Social and cognitive</i>	<i>N</i>	<i>Mean</i>	<i>Mann-Whitney U</i>	<i>Wilcoxon W</i>	<i>Z</i>	<i>Sig.(2-tailed)</i>
Opinion, agree, disagree						
Immediate CF	30	27.50	360.000	825.000	-1.568	.117
Delayed CF	30	33.50				
Question, clarify, request						
Immediate CF	30	29.00	405.000	870.000	-.772	.440
Delayed CF	30	32.00				
Paraphrase, help given						
Immediate CF	30	33.17	370.000	835.000	-1.426	.154
Delayed CF	30	27.83				

Note: $P < 0.05$

Changes in Emotional Engagement

Results concerning participants' affections toward the immediate and delayed CF showed that immediate group had a more positive emotional reaction toward the CF than the delayed group. Table 8 shows the results. The mean score for the immediate CF is 86.26 and 81.80 for delayed CF group. As can be seen, the p value of 0.00 indicates that there was a statistically significance difference between immediate and delayed CF groups in terms of their effect on emotional engagement.

Table 8. Comparison Analysis Results for Engagement Survey

	Group	N	Mean	Std. Deviation	df	f	sig	t
Immediate vs Delayed	Immediate CF	30	86.26	4.55	58	4.92	.000	4.465
	Delayed CF	30	81.80	3.04				

Note: $P < 0.01$

Changes in Complexity Measures

Table 9 depicts the two variables employed to measure the complexity of language use in the participants' oral performances: amount of the use of 'prepositions' and 'conjunctions'. In the case of prepositions, the immediate group outperformed the delayed group (M= 33.50 for immediate; M= 27.50 for delayed group). Statistically speaking, however, there exists no significant difference between immediate and delayed CF in terms of their use of prepositions. Regarding the 'frequency of the use of conjunctions', the immediate group (M= 34.47) outperformed the delayed group (M= 26.53). Likewise, the p value of 0.06 indicated that there was no statistically significance difference between the immediate and delayed groups with regarded to the use of conjunctions.

Table 9. Complexity Measure for Immediate and Delayed Groups

Complexity Measures	N	Mean	Mann-Whitney U	Z	Sig.(2-tailed)
Frequency of the occurrence of prepositions					
Immediate CF	30	33.50	360.000	-1.410	.159
Delayed CF	30	27.50			
Frequency of the use of conjunctions					
Immediate CF	30	34.47	331.000	-1.855	.064
Delayed CF	30	26.53			

Note: $P < 0.5$

Changes in Accuracy Measures

Mean scores of the learners' performance on the effects of immediate and delayed CF on accuracy are presented in Table 10. As can be seen, the p values of 0.21, 0.52, 0.68 revealed no statistically significant differences between the two groups.

Table 10. Accuracy Measures for Immediate and Delayed CF

Accuracy Measures	N	Mean	Mann-Whitney U	Z	Sig.(2-tailed)
Error Free Clauses					
Immediate CF	30	33.08	372.500	-1.232	.218
Delayed CF	30	27.92			
Number of Self- Correction					
Immediate CF	30	29.50	420.000	-.640	.522

Delayed CF	30	31.50			
Target like use of verb tenses					
Immediate CF	30	31.33			
Delayed CF	30	29.67	425.000	-.403	.687

Note: $P < 0.05$

Changes in Fluency Measures

To find answer to the last research question, it was useful to trace the changes, produced by the immediate and delayed CF on fluency which is in part in terms of 'the number of pauses' and 'the number of repetitions'. As can be seen in Table 11, the immediate group obtained a higher score ($M = 32, 31.70$). Thus, the immediate group had more pauses during task performance than the delayed group. Although the difference did not reach statistical significance ($p = 0.46$), delayed CF had a lower number of pauses ($M = 29$) than the immediate group. Similarly, the p value of 0.54 with regard to 'the number of repetitions' revealed no statistical significant difference between the immediate and delayed CF groups. It is important to note that the participants in the immediate CF group paid no attention to the teacher's correction and therefore did not repeat the correct form during task performance.

Table 11. *Fluency Measures for Immediate and Delayed CF*

<i>Fluency Measures</i>	<i>N</i>	<i>Mean</i>	<i>Mann-Whitney U</i>	<i>Z</i>	<i>Sig.(2-tailed)</i>
Number of Pauses					
Immediate CF	30	32.00			
Delayed CF	30	29.00	405.000	-.738	.461
Number of Repetitions					
Immediate CF	30	31.70			
Delayed CF	30	29.30	414.000	-.600	.549

Note: $P < 0.05$

Discussion and Conclusion

This study examined how much progress in learners' CAF, grammar gain and engagement was made as a result of the changes in their CF time. The results of the current study revealed a statistically significant improvement in the two CF groups but did not reveal any significant difference between the two CF conditions with regard to their impact on social and cognitive engagement, grammar gain, as well as on their complexity, accuracy and fluency. On the other hand, in the case of behavioral and emotional engagement, there was a statistically significance difference between the two CF conditions. Next we provide specific answers to the research questions.

The first research question examined whether there was a difference between the pre-test and post-test scores of the immediate and delayed CF conditions and between the two CF groups. The outcomes obtained from the GJT test revealed that there were no statistically significant differences between the two CF conditions, but there was a statistically significant difference between the learners in terms of their patterns of improvement from the pre-test to the post-test.

As it was noted in the reviewed literature, little attention has been paid to the timing of CF, and yet there has been little attempt to distinguish the boundary between immediate and delayed CF. Even though Long (1977) originally defined delayed correction as correction that allows learners to finish their utterances, and immediate correction as correction that interrupts them, contemporary CF research has subsumed these older notions under the label of immediate CF. Contemporary CF literature often refers to delayed CF as a correction that is given sometime after a communicative task (Hunter, 2012; Quinn, 2014).

Following these recommended conventions, a certain period of delay at an optimal length might enable future researchers to close off avenues of hopeless possibilities and to deal with an optimal length of delay which is neither too short nor too long. Undoubtedly, CF that is provided immediately or provided after only a short delay may smother learners' desire to simply use language, unfettered by their over attention to correctness. Closely linked to the immediate CF is the lure of immediate demands of the task or the pressure of knowing that another task will begin shortly, which may cloud students' ability to process new information. Thus, the believing game of distinguishing between immediate and delayed CF may provide future researchers with a means for interpretation and understanding of whether altering the length of delayed CF results in a difference in L2 development and performance.

Finally, and perhaps most importantly, the reason for an increase in scores across the two CF conditions (immediate and delayed) is that the two groups received Form-Focused Instruction (FFI) on the past passive-structure. Goo and Mackey (2013) have maintained that the presence of additional instruction in CF research makes it difficult to clearly attribute results to the effect of CF. They propose that this problem of separating the effects that result from CF from those that result from additional FFI could be avoided by conducting studies that do not include additional FFI. Besides, since the passive structure is less frequently found in natural conversation, it deemed to be necessary to design FFI tasks which would stimulate regular use of the past passive structure upon which CF could be provided when necessary.

The second research question explored whether there was a difference in the L2 learners' engagement that resulted from the immediate and the delayed CF treatment. The results indicated that there was a statistically significant difference between the two CF conditions in the case of behavioral and emotional engagement. On the other hand, no statistically significant differences were found between the two CF conditions in terms of their effect on the L2 learners' social and cognitive engagement.

Participants from the two CF conditions responded to the CF provision with both negative and positive attitudes. Some negative emotional reactions were predictable because, as was evident in the literature, researchers who discourage teachers from using CF (e.g., Krashen, 1982) warn that CF may create negative evidence in learners' mind and undermine their confidence. In contrast, the results obtained from HSSSE revealed that when learners find learning environment exciting and enjoyable, they will pay more attention to the task and yet try harder. Therefore, the difference between these two CF conditions may possibly be due to difference in their corrective functions. Moreover, as Skinner, Furrer, Marchand, and Kindermann (2008) note, learners who are more emotionally engaged show higher levels of behavioral engagement.

In order to see whether there were fluctuations in behavioral engagement in L2 use across the two CF conditions, performance data were measured and then transcribed. The results revealed that the two CF conditions had a statistically significant impact on behavioral engagement. Behavioral engagement, which in this study was marked by the 'total words spoken', did seem to support the development of positive emotions and led to an increase in interest and satisfaction in the immediate CF group compared to the delayed group: Learners with lower level of classroom participation will finally fail to engage in learning tasks and become more bored and frustrated, whereas learners who try hard and hang on in learning tasks will tend to find them increasingly more exciting and fun.

Factors conceptualized as fundamental to cognitive and social engagement revealed no statistically significance difference between the two CF conditions. We found that, in the two CF conditions, student-student interactions seemed to offer more opportunity for establishing high-level cognitive engagement than teacher-student interactions, both in class and small group interactions. The social rules central to the communicative dictogloss task also appeared to play a crucial role in the formation and manifestation of cognitive engagement. As a result, evidence from participants in communicative tasks in general and dictogloss task

in particular, revealed that learning occurs in the midst of communication. This idea is not surprising given the pervasiveness of the communicative approach in EFL classrooms.

The other aim of the study was to examine the spoken complexity and how it can be affected by different CF conditions. In line with previous studies (Farrokhi, Zohrabi & Chehr Azad, 2018; Rahimi & Dastjerdi, 2012), the results did not reveal any significant effects of CF timing on complexity. The results of the descriptive statistics revealed that the immediate CF groups' complexity with regard to the 'frequency of the use of prepositions' and 'use of conjunctions' was lower than that of delayed CF. However, considering the inferential statistics, there were insignificant differences between the two CF groups in either occurrences.

The logical explanation for these results, according to Farrokhi et al. (2018) can be based on Schmidt's (1995) distinction between low and high levels of awareness and the argument that while noticing is necessary for acquisition, understating results in deeper learning.

Compared to other aims of consciousness, consciousness at the level of understanding, a higher level of awareness, has had a relative influence on an analysis of an object of consciousness in meaningful learning. Therefore, it can be suggested that since "explicit correction helps learners develop awareness at both levels, noticing and understanding, it can be a better candidate for the promotion of the second language learning" (Farrokhi et al, 2018, p. 138). Besides, the participants in this study may not still have reached an adequately advanced linguistic level in their L2 development to begin displaying a clear attention to complexity (Skehan, 2009).

Finally, in contrast to previous studies (e.g. Farrokhi et al, 2018; Rahimi & Dastjerdi, 2012), the analysis reported on the effects of the two CF provisions on the fluency and accuracy revealed no statistically significant difference between the CF groups. One possible reason for this lack of difference may be learners' pair work opportunities to bring together their linguistic experience and co-construct L2 language knowledge. Swain and Lapkin (1998) indicate that requests for validation of language form or language choice draw learners' attention to a specific language form. This possibly makes learners more receptive, confident and hence more likely to notice the feedback received in small group interactions which, in turn, may affect their accuracy and fluency.

The findings of the present study call for more flexibility on the part of the teachers in providing either immediate or delayed CF whenever appropriate and without being concerned

about the adverse effects of feedback. Such flexibility is academically beneficial for teachers, especially when they need to combine instruction and CF also need to wait until after some practice time before CF provision. However, accuracy may possibly be a similarly essential element of the effective completion of the task. For instance, automaticity and accuracy are necessary for those L2 students who are learning how to use rhetorical devices in public speaking classes which may, in turn, affect their engagement.

Study Limitation

The limitation of the study indicates the need for more research which includes longer interventions. The longer one might bring in its differences between immediate and delayed CF treatments in a more reliable way.

References

- Butler, D. L., & Winne, P. H. (1995). Feedback and self-regulated learning: A theoretical synthesis. *Review of Educational Research*, 65(3), 245-281.
- Cohen, A. D. (1987). Student processing of feedback on their compositions. In A.L. Wenden and J. Robin (eds) *Learner Strategies in Language Learning*, 57-69. Englewood Cliffs, NJ: Prentice-Hall.
- Ellis, R., (1994). *The study of second language acquisition*. Oxford University.
- Ellis, R. (2003). *Task-based language learning and teaching*. Oxford University Press.
- Ellis, R. (2010). Epilogue: A framework for investigating oral and written corrective feedback. *Studies in Second Language Acquisition*, 32(2), 335-349.
- Ellis, R., & Barkhuizen, G. P. (2005). *Analyzing learner language*. Oxford: Oxford University Press.
- Farmani, R., Akbari, O., & Ghanizadeh, A. (2017). The Impact of Immediate and Delayed Error Correction on Iranian EFL Learner's Motivation. *European Journal of Foreign Language Teaching*, 2(3), 76-88.
- Farrokhi, F., Zohrabi, M., & Chehr Azad, M. H. (2018). Corrective feedback and Iranian EFL learners' spoken complexity and accuracy. *Teaching English Language*, 12(2), 117-143.
- Ferris, D. R. (1995). Student reactions to teacher response in multiple-draft composition classrooms. *TESOL Quarterly*, 29(1), 33-53.
- Goo, J., & Mackey, A. (2013). The case against the case against recasts. *Studies in Second Language Acquisition*, 35(1), 127-165.

- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81-112.
- Hedgcock, J., & Lefkowitz, N. (1996). Some input on input: Two analyses of student response to expert feedback in L2 writing. *The Modern Language Journal*, 80(3), 287-308.
- Hendrickson, J. (1978). Error correction in foreign language teaching: Recent theory, research, and practice. *The Modern Language Journal*, 62(8), 387-98.
- Housen, A., & Kuiken, F. (2009). Complexity, accuracy, and fluency in second language acquisition. *Applied Linguistics*, 30(4), 461-473.
- Housen, A., Kuiken, F., & Vedder, I. (Eds.). (2012). *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA* (Vol. 32). John Benjamins Publishing.
- Hunter, J. (2012). Small talk: Developing fluency, accuracy, and complexity in speaking. *ELT Journal*, 66 (1), 30-41.
- Hunter, J.D., (2011). *A multi-method investigation of the effectiveness and utility of delayed corrective feedback in second language oral production*. Unpublished Ph.D. Dissertation (Module 3), University of Birmingham, UK.
- Krashen, S. D. (1982). *Principles and practice in second language learning*. NY: Pergamon.
- LaPierre, D. (1995). *Language output in a cooperative learning setting: Determining its effects on second language learning*. Unpublished MA thesis, University of Toronto.
- Larsen-Freeman, D. (1995). On the teaching and learning of grammar: Challenging the myths. In F. Eckman, D. Highland, P. Lee, J. Mileham & R. Rukowski Weber (Eds.), *Second language acquisition theory and pedagogy*. Mahwah, NJ: Lawrence Erlbaum.
- Lee, E. J. E. (2013). Corrective feedback preferences and learner repair among advanced ESL students. *System*, 41(2), 217-230.
- Li, S. (2010). The effectiveness of corrective feedback in SLA: A meta-analysis. *Language Learning*, 60(2), 309-365.
- Li, S., Zhu, Y., & Ellis, R. (2016). The effects of the timing of corrective feedback on the acquisition of a new linguistic structure. *The Modern Language Journal*, 100(1), 276-295.
- Loewen, S., & Nabei, T. (2007). Measuring the effects of oral corrective feedback on L2 knowledge. *Conversational interaction in second language acquisition: A collection of empirical studies*, 361-377.

- Long, M. H. (1977). Teacher feedback on learner error: Mapping cognitions in H. D. Brown, C. A. Yorio, & R. H. Crymes (Eds.), *On TESOL '77* (pp. 278-293). Washington: TESOL. 278-294.
- Lyster, R., & Saito, K. (2010). Oral feedback in classroom SLA: A meta-analysis. *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.
- MacIntyre, P. D., Dörnyei, Z., Clément, R., & Noels, K. A. (1998). Conceptualizing willingness to communicate in a L2: A situational model of L2 confidence and affiliation. *The Modern Language Journal*, 82(4), 545-562.
- Philp, J., & Duchesne, S. (2016). Exploring engagement in tasks in the language classroom. *Annual Review of Applied Linguistics*, 36, 50-72.
- Quinn, P. (2014). *Delayed versus immediate corrective feedback on orally produced passive errors in English*. (Unpublished doctoral dissertation). University of Toronto, Toronto.
- Rahimi, A., & Dastjerdi, H. V. (2012). Impact of immediate and delayed error correction on EFL learners' oral production: CAF. *Mediterranean Journal of Social Sciences*, 3(1), 45-54.
- Rocca, K. A. (2008). Participation in the college classroom: The impact of instructor immediacy and verbal aggression. *The Journal of Classroom Interaction*, 22-33.
- Salend, S. J. (2001). *Creating inclusive classrooms: Effective and reflective practices* (4th ed.). Columbus, OH: Merrill/Prentice Hall.
- Saslow, J. M., Ascher, A., Morsberger, R. E., Laporte, P., Long, W. P., & Ruzicka, D. (2011). *Top notch*. Pearson Education.
- Schmidt, R. W. (1995). Consciousness and foreign language learning: A tutorial in the role of attention and awareness in language learning. In R. Schmidt (Ed.) *Attention and awareness in language foreign language learning*, (pp. 1-64). Honolulu, HI: University of Hawaii Press.
- Shabani, K., & Safari, F. (2016). Immediate vs delayed correction feedback (CF) and accuracy of oral production: The role of anxiety. *Theory and Practice in Language Studies*, 6(11), 2222-2230.
- Sheen, Y. (2012). *The timing of corrective feedback and L2 learning*. Paper presentation at the Second Language Research Forum, Carnegie Mellon University, U.S.A.

- Skinner, E., Furrer, C., Marchand, G., & Kindermann, T. (2008). Engagement and disaffection in the classroom: Part of a larger motivational dynamic? *Journal of educational psychology*, *100*(4), 765.
- Skehan, P. (2009). Modelling second language performance: Integrating complexity, accuracy, fluency, and lexis. *Applied Linguistics*, *30*(4), 510-532.
- Stroud, R. (2015). *Learner Classroom Engagement: Definition, Measurement and Data Usage*. Paper presentation at the Asian Conference on Psychology and the Behavioral Sciences. Kwansai Gakuin University, Japan.
- Stroud, R. (2017). The impact of task performance scoring and tracking on second language engagement. *System*, *69*, 121-132.
- Swain, M., & Lapkin, S. (1998). Interaction and second language learning: Two adolescent French immersion students working together. *The Modern Language Journal*, *82*(3), 320-337.
- Varnosfadrani, A. D. (2006). *A comparison of the effects of implicit/explicit and immediate/delayed corrective feedback on learners' performance in tailor-made tests*. Unpublished PhD thesis, University of Auckland.
- Yazzie-Mintz, E. (2007). *Voices of Students on Engagement: A Report on the 2006 High School Survey of Student Engagement*. Center for Evaluation and Education Policy, Indiana University.

