The Effect of Written Corrective Feedback on EFL Learners’ Performance after Collaborative Output

Shirin Abadikhah*
University of Mazandaran, Babolsar, Iran
Email: sh_abadikhah@yahoo.com

Azadeh Ashoori
Iran Language Institute, Sari, Iran
Email: azadehashoori@yahoo.com

Abstract—The purpose of the present study is to investigate the effect of written corrective feedback on EFL learners’ performance after completing a set of output activities. The study involved two groups of learners (24 male students) at intermediate level. The first group, consisting of six pairs, worked on four output activities (text editing, composition, transformation and substitution). The second group received written feedback after completing the same activities. Twelve tailor-made dyad-specific tests were constructed and administered to the pairs of learners. Results revealed that the participants who received written corrective feedback after completing the activities outperformed those who did not receive written feedback. The study confirms the view that providing an opportunity for interaction after receiving feedback triggers noticing the gaps between the learners’ interlanguage and target language, which facilitates L2 development.

Index Terms—corrective feedback, output activities, collaborative dialogue

I. INTRODUCTION

Numerous studies have been conducted on the types of feedback given to learners and their effectiveness in learning various linguistic features (Carroll, 1996; Carroll & Swain, 1993; Dekeyser, 1993; Doughty, 1994; Spada & Lightbown, 1993). Although early studies raised considerable doubts as to the efficacy of feedback (e.g. Chaudron, 1977; Chun et al., 1982), recent studies provide substantial evidence on its effectiveness in SLA (Carroll & Swain, 1993; Ellis et al., 2001; Ferris, 1997). Studies in the 1990s denote that providing feedback to students on both content and form may improve their learning and result in the development of their interlanguage system.

From a theoretical perspective, a number of proposals have been made in support of the beneficiary role of corrective feedback in SLA. Schmidt's (1990, 2001) Noticing Hypothesis states that noticing is a prerequisite for L2 learning and learners must pay conscious attention to input in order for learning to occur. The proponents of this hypothesis support the facilitative role of corrective feedback in drawing learners' attention to form. They believe that corrective feedback provides a stimulus for learners to identify the gap or mismatch between their interlanguage and the target language. Swain's (1985, 1995) Output Hypothesis also assigns a primary role for corrective feedback. While producing output, learners may experience difficulties in conveying their intended meaning. By providing corrective feedback, in fact, the relevant input would be immediately available and the learners may process language with focused attention. Finally, Long's (1996) extended version of Interaction Hypothesis ascertain the facilitative role of corrective feedback in language learning. Corrective feedback can provide learners with direct and indirect information about the grammaticality of the utterances. Long (1996) argues that "negoiation for meaning, and esp. negotiation work that triggers interactional adjustments by the NS or more competent interlocutor, facilitates acquisition because it connects input, internal learner capacities, particularly selective attention, and output in productive ways" (pp. 451-452).

The most considerable classfication of corrective feedback includes direct and indirect feedback. Direct corrective feedback occurs when the learners are overtly informed of the existence of an error and provided with the target-like reformulation by their teachers. This type of feedback involves either explicit correction or metalinguistic feedback (Ellis et al., 2006). In explicit correction, the teacher clearly states that what the student has produced is erroneous, whereas in metalinguistic feedback, the teacher provides students with information, comments and questions with respect to the well-formedness of their production (Lyster & Ranta, 1997). In Indirect feedback such as highlighting, underlying or coding, the teacher makes students understand that they have made an error.

* Corresponding Author: Shirin Abadikhah, Department of English Language and Literature, Faculty of Humanities and Social Sciences, University of Mazandaran, P.O. Box: 416, Babolsar, Iran. e-mail:<sh_abadikhah@yahoo.com>
From a theoretical standpoint, learners must pass through four crucial stages in order to learn a second language. According to Garcia-Mayo and Pica (2000), learners must (a) be provided with meaningful and comprehensible input, (b) pay attention to the form and meaning of the input, (c) produce the target language and finally (d) receive feedback in order to modify their production towards greater accuracy, comprehensibility and appropriateness. There has been little emphasis on the final two stages of L2 learning in the Iranian EFL context. The premise of the present study evolves around the role of feedback in focusing learners' attention to their output after collaboration.

II. LITERATURE REVIEW

Swain (1995) emphasized the role of modified output in second language acquisition. In fact, modified output is the representation of the leading edge of a learners' interlanguage. She suggested that useful feedback from peers could provide opportunities for modified output. More specifically, corrective feedback, defined as "any indication to the learners that their use of the target language is incorrect" (Lightbown & Spada, 1999, p. 171) can be considered as a kind of modified input, which may lead to modified output on the part of the learner. As learners struggle to test their hypotheses, the external feedback (implicit or explicit) provided by their teachers or classmates can help them notice the problems in their language. When there is no external feedback, learners do not have anything to test their hypotheses against (Swain, 1995). Swain (1997) stressed the significance of providing feedback to learners' collaborative work and argued that when learners complete the task in collaboration, they tend to stick to the knowledge they have constructed together and transfer their knowledge (correct or incorrect) to the following posttest. If they reach correct resolution, they tend to be accurate in their responses and if they do not, they show tendency to choose inaccurate responses (Swain, 1998, pp. 78-79). Thus, it is clear that without feedback, learners may remember their incorrect solutions to their linguistic problems and may transfer them to the new learning situation. In this case, learning also happens but not learning the correct structure.

Earlier, in an experimental study, Carroll, Roberge and Swain (1992) found that learners who received feedback outperformed those who did not receive feedback in the acquisition of the target linguistic forms. Carroll and Swain (1993) further examined the relative effectiveness of various types of feedback on the acquisition of English dative alternation by ESL learners. They found that the provision of negative feedback facilitated learning the selected features. That is, all types of feedback led to learning, but the participants receiving the most explicit (or direct) form of feedback outperformed the other three groups.

Ferris (1997) argues that giving feedback to students on both content and form with underlined errors in their essays might improve their writing. She states that when her learners were given grammatical comments, they revised their writing more efficiently. In an empirical study, Ferris (2002) found that error correction feedback was beneficial because it improved learners' accuracy in the short term. She realised that her students gained positive attitudes about the effect of feedback, which encouraged them to be independent self-editors. Ferris and Roberts (2001) also found that beginner L2 students benefited from being corrected by their teacher. Without feedback, students hardly noticed their own errors, and as a result, frustration occurred. They believe that indirect feedback increases the learners' reflection and attention to the erroneous forms and similarly involves them in guided learning, which may cause long-term retention. Nassaji and Swain (2000) also argued that when input is not sufficient to help learners use correct form and there is no consistency between what they say and what is normally said in the target language, negative feedback can be effective for language learners to prevent them from overgeneralizing from their L1 into L2. They further emphasized that the type of negative feedback is very crucial in its usefulness. That is, the more the feedback is direct, the more it draws learners' attention to their error. Chandler (2003) also states that direct feedback is more effective for learning because it does not confuse the students, gives them more information and provides them with the opportunity to receive immediate feedback, consequently preventing them from making wrong hypotheses. Considering the growing consensus among the majority of scholars on the significant role of direct corrective feedback in SLA, the researchers assumed this type of feedback to be most appropriate for the learners in the context of this study.

Although numerous studies have been conducted on corrective feedback, the bulk of the studies focused on the product of learning, few studies have delved into provision of feedback following collaboration on a set of output tasks. The present study seeks to answer the following research questions:

1) Does collaborative output tasks affect EFL learners' subsequent performance?
2) Do learners who receive corrective feedback after collaboration on a set of output activities outperform those who do not receive corrective feedback?

III. RESEARCH METHODOLOGY

A. Participants

Twenty-four male students within the age range of 15 to 26 were chosen for this research. They were invited from a high intermediate level class at the Iran Language Institute (ILI) in Sari, Iran. All of the students were studying English in the foreign language setting with Persian as their mother tongue. These students formed two groups of six pairs (12 students in each group). In order to increase their interaction, they were free to choose their partners. Each pair in the
first group (non-feedback) was asked to attend their classes in a specified time of the first week. The following week was devoted to the second group (feedback).

B. Instrumentation

For conducting the present research, four output activities (text editing, composition, transformation and substitution) were utilized. These activities were taken from the high intermediate ILI English series; therefore, the items were consistent with the participants' proficiency level. The students were encouraged to help each other, correct themselves and one another during the completion of the activities and pay attention to form when errors are made.

The text-editing task required the students to repair a text in order to improve its accuracy. The text was a letter to a friend with 182 words in length and contained 14 mistakes on the use of gerund. Working in pairs, the students were asked to identify the errors. The second task was a composition for which the students were asked to write a paragraph of 100 words in length together on their own favorite topics. They were told that they could not use a dictionary or have researchers’ aid during the task. Before starting writing, the students were given time to select their topics. They were also free to write in past, present or future tense. In addition to these tasks, two drills including transformation and substitution, representing mechanical drills in their textbooks, were employed. The transformation drill consisted of 10 active sentences that should be changed into passive voice. The substitution drill consisted of 16 words that should be substituted in selected sentence patterns (see Appendix A for a sample task). There was no time limit for doing these tasks and drills. Completing the whole set of activities varied from pair to pair and took about 40 to 90 minutes. To have a good voice quality, an MP3 player was used to record the conversation between the learners.

C. Procedure for Data Collection

One week before starting the main study, one of the researchers devoted some time to explain the aim and procedure of the study to the students. The participants who were interested in the study were asked to sign a consent form to show their willingness. Twenty-seven students volunteered to participate in the study. Having chosen their partners, each pair was assigned a date and informed of the time of the attendance. The research was carried out in fourteen sessions during four weeks. One of the researchers was present during all these sessions. Before starting each session, she explained how to complete the assignments for each dyad. To encourage joint production, each pair was given only one copy of the activities. During the first week, the first group (non-feedback) completed the activities without receiving any written feedback. On the second week, the next group (feedback) was given the same tasks and drills followed by written feedback by the researchers. The written feedback consisted of explicit provision of the correct forms with no metalinguistic information about the well-formedness of the learners’ utterances (see Appendix B for sample feedback).

Ten days after the last session, all pairs were given tailor-made dyad-specific tests in one session. Tailor-made test was used by LaPierre (1994, cited in Swain, 1995) in a pilot study examining the output hypothesis and later by Williams (2001). It is based on the dialogues of pairs of students when they are talking about the language (Swain, 1998). The test aims to assess the students’ knowledge which may be constructed through metatalk. This kind of test was created to see whether the participants correctly use the items which were focused on during their collaboration (see Appendix C for sample test). To design the tests, the tape-recorded interaction of the participants was transcribed and analyzed in terms of linguistic features which has been focused on. The data were checked for the occurrence of language-related episodes (LREs) by the two researchers. An LRE is defined as any part of the students’ conversation in which they talk about the language (their problems) they are producing, question their language use, or self- and other-correct (Swain, 1997). Agreement rate for identification of LREs between the raters was 95 percent and those four disagreements between the two coders were resolved by discussion. The tailor-made tests were constructed from the focused items in all tasks and drills. The number of items in the tests was the same for the two members of each pair but varied from pair to pair. To estimate the validity of these tests (constructed for 12 pairs) two ILI teachers were provided with 10% of the transcripts along with the testing items on the linguistic episodes focused during the students’ interaction. The aim was to determine to what extent the test items measured the learners’ ability to produce linguistic forms they talked about during their collaboration. They were asked to rate the items in the tests on two scales (a) corresponded to appropriate and (b) corresponded to inappropriate. The result from the rating showed considerable agreement between the raters (95.55%).

The scoring procedure was very strict and for every correct answer, one score was considered and incorrect answers were scored zero. For further clarification, the design of the study is presented briefly in the following table.

<table>
<thead>
<tr>
<th>Sessions</th>
<th>Activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Week 1</td>
<td>NFG completed the activities and were tape-recorded separately</td>
</tr>
<tr>
<td>Week 2</td>
<td>FG completed the assignments, received feedback and were tape-recorded separately</td>
</tr>
<tr>
<td>Week 3</td>
<td>Listening to their interaction, transcribing and preparing tests</td>
</tr>
<tr>
<td>Week 4</td>
<td>Administering dyad-specific tests for 12 pairs</td>
</tr>
</tbody>
</table>

**TABLE 1**

THE DESIGN OF THE STUDY

FG= Feedback group; NFG= Non-feedback group

D. Data Analysis
This study examines whether collaboration and provision of feedback after collaboration would affect EFL learners' performance. To achieve this goal, the recorded pair-talk for both groups was analyzed in terms of language-related episodes (LREs). Following Swain’s (1997) definition, self-corrections were also regarded as episodes. It is worth stating that immediate self-corrections were not considered as LREs because it was difficult to distinguish between the self-corrections which were as a result of reflection on language and those which were the performance slip without much reflection on form. Therefore, those self-corrections in which the students explained or reflected on the linguistic forms for some time were regarded as LREs. The total time spent on the completion of the activities was 538 minutes, during which 316 LREs were produced. The identified LREs were further classified according to the type of resolution—i.e., the way the students resolved their problems. Following Swain (1995, 1998), LREs were classified into three groups: a) correctly resolved LREs; b) incorrectly resolved LREs and c) unresolved. As the term suggests, correctly solved LREs are those episodes in which the pair members reach the correct answer after collaboration. Sometimes one student in a pair made an error which was corrected by the other member and sometimes both of the students reached the correct answer after a short discussion. With respect to the incorrectly solved LREs, the pairs solved their problems together but they did not obtain target-like forms. Finally, unresolved LREs occurred when the learners did not reach a resolution and continued their conversation.

IV. FINDINGS AND DISCUSSION

A. Evidence from Product Data

Our research questions concerned whether collaboration and written feedback after collaboration affect EFL learners' subsequent performance. To address this issue, a group of EFL learners consisting of 6 pairs received written feedback immediately after completing the activities. The correct responses to the assignments were prepared before the research sessions started. It is worth mentioning that the feedback on composition was not given to the students immediately after completion since it needed some time to correct their compositions. The number of LREs produced in the four activities by the two groups of participants is presented in Table 2.

<table>
<thead>
<tr>
<th>THE NUMBER OF LREs PRODUCED BY 12 PAIRS IN THE FOUR ACTIVITIES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Text- editing</td>
</tr>
<tr>
<td>133</td>
</tr>
</tbody>
</table>

After identification of LREs, they were coded based on their type of resolution. The total number of LREs in the transcripts was 316, which consisted of 189 correctly solved, 113 incorrectly resolved and 14 unresolved episodes. The frequency and percentage of each LRE type are provided in Table 3.

<table>
<thead>
<tr>
<th>FREQUENCY AND PERCENTAGE OF LRE RESOLUTION TYPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correctly-solved</td>
</tr>
<tr>
<td>No of LREs</td>
</tr>
<tr>
<td>% of LREs</td>
</tr>
</tbody>
</table>

As was stated, a tailor made test was given to each individual dyad in the non-feedback group (who did not receive written feedback) and in the feedback group (who received written feedback). Unlike the activities, the tests were completed individually by each member of the dyads. This test was constructed on the basis of each pair’s correctly solved episodes. Having administered the test, the researchers went through the students' papers to see how many items were answered correctly out of the total number of questions in the test. To answer the first research question, the obtained data from the dyads in the first group who did not receive written feedback were used. That is, the learners' correctly solved LREs were compared with their scores in the tailor-made test. The summary of the results is presented in the following table.

<table>
<thead>
<tr>
<th>SUMMARY OF THE RESULTS OF COMPARISON OF THE MEANS (NON-FEEDBACK GROUP)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correctly solved LREs</td>
</tr>
<tr>
<td>N</td>
</tr>
<tr>
<td>Mean</td>
</tr>
<tr>
<td>SD</td>
</tr>
<tr>
<td>t</td>
</tr>
<tr>
<td>df</td>
</tr>
<tr>
<td>p</td>
</tr>
</tbody>
</table>

The table reveals that the learners’ mean scores declined in the testing session. To tackle the research question, we employed a matched t-test and compared the mean scores of the students’ correctly solved LREs in during task completion with their correct responses to the tailor made test. The result showed that the mean scores of the two sets of data were statistically different, (t= 2.68, df=11, p= 0.01), as the p value came out to be lower than 0.05. This suggests that the performance of the non-feedback group in the tasks significantly differed from their performance in the tests. That is, the pairs’ mean score in the correctly solved LREs was significantly higher than their mean score in the subsequent performance. The result may imply that the students' co-constructed knowledge during collaboration (as was revealed in their correctly solved LREs) was not fully transferred to their subsequent performance in the tailor-made test.
The second research question concerns whether the two groups of learners (NFG and FG) differed in their performance after completing the assignments in pairs. To find an answer for this question, we compared the mean scores of the two groups in the tailor-made test. The summary of this analysis is presented in the following table.

<table>
<thead>
<tr>
<th>Groups</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>t value</th>
<th>df</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>NFG</td>
<td>12</td>
<td>68.41</td>
<td>18.66</td>
<td>2.36</td>
<td>22</td>
<td>0.028</td>
</tr>
<tr>
<td>FG</td>
<td>12</td>
<td>73.33</td>
<td>10.47</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In Table 5, the results of the t-test analysis shows a statistically significant difference between the feedback and nonfeedback groups' performance in the test (t =2.36, df=22; p=0.028). This finding suggests that working in pairs may not guarantee better performance in future and in addition to peer feedback, an external feedback from a more knowledgeable interactant is needed to confirm or disconfirm the learners’ constructed hypotheses about the target language.

B. Evidence from the Process Data

Receiving written feedback may facilitate the recognition of problems and may encourage learners to 'notice' the gaps in their interlanguage. The evidence for this comes from the dialogues of the learners after receiving feedback. In the following extracts, it is shown that how written feedback made them aware of their linguistic problems, or gaps, as Swain (1998) has pointed out.

_Extract 1_
1: We look forward to seeing? Seeing is correct? To is a preposition?
2: Yes, because 'seeing' is right.
1: I didn't pay attention to this.

Extract 1 was taken from the students’ conversation after receiving feedback on the text-editing task. They compared their own responses with the correct responses provided. S1 knew that the verb assumes “ing” form after preposition, however he was not sure about this case. He asked S2 for help. In the end, he expressed that he did not consider this particular case. It seems that provision of feedback focused their attention to the problem.

_Extract 2_
1: A vacation has been promised.
2: Has been? Not have been?
1: Oh, yes. Vacation. Vacation is singular.
2: Yes. Yes.

This conversation occurred after two members of a pair received feedback on transformation drill. Their interaction evolved around the use of verb 'have' or 'has' for the subject 'vacation'. Having compared their answer with the feedback, they noticed that the verb 'have' does not agree with the subject 'vacation'. S1 did not get the point at first but after S2 repeated the verbs, he attended to the problem and noticed the difference. Following that, he gave the metalinguistic explanation by expressing the rule.

There are numerous examples in the pairs’ conversation indicating that they had the chance to compare their co-constructed answers with the correct responses. This is in accordance with Swain and Lapkin’s (2001) statement that the comparison stage between their own output and the reformulated version promotes learners' noticing. This finding lends further support to Schmidt's (1990) noticing hypothesis, which emphasizes the role of feedback in drawing learners' attention to form and recognizing the gap between their linguistic knowledge and the target forms. As Swain (1995) puts it, this form of noticing, which is one of the functions of output, causes a cognitive process that is helpful in the process of SLA.

V. CONCLUSION

The present study was designed to explore whether collaborative output and written corrective feedback have any effect on the performance of EFL learners after pair-work activity. The findings indicated that the students who received written feedback on their productions outperformed those who did not receive feedback. By receiving feedback, the participants had the opportunity to engage in 'cognitive comparisons' between their own responses and the correct answers. The transcripts revealed that receiving written feedback after completing the activities in pairs triggered noticing the gaps between their output (or interlanguage) and the target forms. The feedback may have provided a condition in which learners could test their generated hypotheses and were able to confirm or disconfirm their grammar as manifested in their output (Doughty, 2001). This finding is consistent with Swain’s (1985,1995) proposal that external feedback may enable learners replace the incorrect hypotheses about the target language with the correct ones.

While the findings of this study suggest effective evidence for the practice of written feedback on EFL students' pair work, the results should be interpreted within the present context with the limited number of participants, types of tasks and drills. Accordingly, future research is needed to replicate the study using more participants with different
proficiency levels, ages and gender as well as various tasks and activities in other modalities. The present study provides language teachers with information about the effectiveness of collaborative dialogue followed by written feedback. Teachers can motivate learners to see their classes in a positive light. Providing an opportunity for learners to participate in collaborative tasks and similarly interaction after receiving feedback may give them a strong sense of motivation. The students may also show positive attitudes towards receiving teacher feedback and improving their writing.

APPENDIX A: TEXT-EDITING TASK

There are fourteen missing gerunds in this letter. Find and correct them.

Dear Adam

I've been here for three days and am having a great time, but I can't help wishing you were here too. Tell your boss I'm really angry at him. He is not letting you take any vacation time. Believe it or not, the first night I missed hearing all the city noises, but I haven't really had any trouble getting used to the peace and quiet since then. Everything is all so relaxed here – there is no rush around or writing things down in your Daily Planner. Getting out of New York City was definitely what I needed, even if it's only for two weeks. The ranch has lots of activities-horseback ride, river raft and hike. Tomorrow we're all going to Taos Pueblo to watch some weave being done and to see some Native American dance, which is great because I'm really interested in learning more about Native American culture. And I'm looking forward to seeing The Magic Flute at the Santa Fe Opera on Saturday. I'll write again in a day or two. Miss you lots.

Love,
Louise

APPENDIX B: FEEDBACK ON TEXT-EDITING TASK

There are fourteen missing gerunds in this letter. Find and correct them.

Dear Adam

I've been here for three days and am having a great time, but I can't help wishing you were here too. Tell your boss I'm really angry at him. He is not letting you take any vacation time. Believe it or not, the first night I missed hearing all the city noises, but I haven't really had any trouble getting used to the peace and quiet since then. Everything is all so relaxing here – there is no rushing around or writing things down in your Daily Planner. Getting out of New York City was definitely what I needed, even if it's only for two weeks. The ranch has lots of activities-horseback riding, river rafting and hiking. Tomorrow we're all going to Taos Pueblo to watch some weave being done and to see some Native American dancing, which is great because I'm really interested in learning more about Native American culture. And I'm looking forward to seeing The Magic Flute at the Santa Fe Opera on Saturday. I'll write again in a day or two. Miss you lots.

Love,
Louise

APPENDIX C: SAMPLE TEST ITEMS DESIGNED FOR DYAD 2 (FEEDBACK GROUP)

Choose the best answer.
1. The movie was so sad that they couldn't help ......................
   a) cry                     b) to cry                  c) crying                  d) cried
2. ...................... experience is important, so we all told her to keep trying.
   a) Getting                b) To get                  c) She's getting           d) She was getting
3. They told us that they very much looked forward to ...................... them.
   a) visit                        b) visiting                c) visited                  d) have visited
4. We all celebrated .................. opening Sue's kitchen.
   a) she                           b) hers                    c) she was                  d) her
5. It's no good ...................... me about your lost cat.
   a) to tell                      b) tell                   c) him tell                  d) telling
6. She objected to .................. to his birthday party.
   a) go                           b) going                   c) went                     d) gone
7. If you aren't sure of its meaning, you can look .............. in a dictionary.
   a) up it                       b) at it up                c) it up                    d) for it up
8. The children were asleep. They were very tired, so I didn't ..............
   a) wake up them                b) wake him up             c) wake them up            d) wake up him
9. When we went to the beach, many people .................. volleyball.
   a) playing                      b) was playing            c) were playing             d) are playing

REFERENCES
Shirin Abadikhah is assistant professor of TEFL at the University of Mazandaran, Iran. She received her PhD in applied linguistics from the University of Edinburgh, UK, her M.A. in TEFL from Tarbiat Modarres University and her BA in English Translation from Allame Tabatabaee University, Iran. She teaches EFL courses at undergraduate and graduate levels at Mazandaran University. Her main research interests include applied linguistics, second language acquisition.
Azadeh Ashoori received her MA in TEFL from Mazandaran University in 2010. She has been teaching English to young adults at the Iran Language Institute (Sari) since 2007. Her areas of interest are pair-work interaction, collaborative dialogue and teaching English to young learners.