The Effects of Written Corrective Feedback Techniques on EFL Students’ Control over Grammatical Construction of Their Written English

Ataollah Maleki
Zanjan University of Medical Sciences, Iran

Elham Eslami
Tehran Institute of Technology, Iran

Abstract—There has been a controversy over the usefulness of the written corrective feedback (WCF) on the accuracy of ESL/EFL learners’ writing – especially after Truscott’s assertion that grammar correction is pointless and harmful (1996). This study presents the findings of an investigation of the impact of WCF on 90 intermediate Iranian EFL students. The participants were separated into three groups; then they randomly received direct, indirect or no correction feedback. They created three pieces of writing, pre-test, immediate post-test and delayed post-test. Simple past tense errors were brought into focus in the feedback. The results showed that the recipients of WCF achieved more than those in the control group—suggesting the effectiveness of both kinds of WCF. Therefore, the provision of WCF should be regarded as a potentially valuable technique in instructing writing to EFL learners.

Index Terms—grammar correction, corrective feedback, meta-linguistic explanation, direct feedback, indirect feedback

I. INTRODUCTION

Improving students’ writing accuracy is an essential factor in effective writing. Effectiveness of a piece of writing will be determined in part by its accuracy. This is the reason why grammar correction has received so much attention in the recent decades. There exist two opposite perspectives regarding grammar correction which is identified as a potential focus-on-form instrument. These different perspectives stem from different language teaching methodologies which impose the utilization of different grammar correction feedback techniques and strategies. The proponents of grammar correction claim that using corrective feedback (CF) technique could significantly improve students’ writing accuracy, whereas the opponents consider it as ineffective and even harmful.

To written corrective feedback (WCF) – which is the focus of this study – Truscott (1996) paid special attention. He pointed to the fact that there was no sufficient research in favor of grammar correction. He also claimed that grammar correction is neither helpful nor effective, and worse it is harmful. This claim led to a controversy among the specialists in the field – which in turn yielded a growing body of research.

A number of studies examined the research results for or against grammar correction and concluded that earlier research findings failed to provide clear evidence whether WCF helps learners improve linguistic accuracy (Ferris, 1999; Guenette, 2007). Many researchers have done studies to clear the issue; however, the experts in the field are still unable to reach a conclusion. As extensive reviews of the available empirical research show (see Hyland & Hyland, 2006; Goldstein, 2004, 2005) the findings about the efficacy of WCF are mixed and thus inconclusive.

Different studies have been done in the domain of written corrective feedback. Many of them involve investigating the efficacy of WCF. Some of them, however, do not include no-feedback control group (Carroll, et al., 1992; Chandler 2000, 2003; Ferris, 1995, 1997, 2006; Lalande, 1982); those which include a control group do not examine the students delayed performance on new pieces of writing, rather they just considered the students’ revisions or immediate post-tests (Fathman & Whalley, 1990; Ashwell, 2000; Ferris & Roberts, 2001). In contrast, the main objective of the current research is to explore the effect of WCF on the participants’ delayed post-test by including a no-feedback control group.

Most of the studies on WCF make a distinction between two kinds of corrective feedback, namely direct CF and indirect CF (Bates, et al., 1993; Ferris, 1995; Ferris & Hedgescock, 1998; Hendrickson, 1978, 1980; Lalande, 1982; Walz, 1982). In the case of direct CF the students are provided with the correct form (Ellis, 2009) the teachers cross out an unnecessary word, insert a missing word, and write the correct form. This type of feedback is desirable for low-level-of-proficiency students who are unable to self-correct, and can not provide the correct form. However, the learners perform
the least processing and thus it does not contribute to long-term learning (Ellis, 2009). However, Sheen (2007) indicates that direct CF can be beneficial for learning only some specific grammatical features.

Indirect feedback occurs when the students are informed in some way that an error exists but are not provided with the correct form, thus placing the burden of spotting the erroneous forms on students. The experts in the field argue that indirect feedback is superior for most students, because it involves them in “guided learning and problem solving” (Lalande, 1982), focusing their attention to linguistic forms that may lead to long-term learning (Ferris & Roberts, 2001; James, 1998; Reid, 1998). However, the findings of different studies which have focused on the difference between direct and indirect CF are very mixed. Some studies (Ferris & Helt, 2000; Lalande, 1982) claim that indirect feedback enables students to correct their errors, however, some suggest the opposite (Chandler, 2003), and others (Robb, et al., 1986; Frantzen, 1995) found no difference.

All in all, although, according to research, students strongly prefer the direct red pen technique (underlining and description), overall results show that it is not sufficient to provide CF on students’ writing and Some mini-lessons or workshops are essential to enable students to self-edit.

Moreover, as Reid (1998) argued, because of the existence of dissimilar motivations for L2 writing and different perceptions of English learning, the kinds of appropriate grammar correction for EFL and/or international students may be different from what is beneficial for immigrant students who are mainly “ear learners” and their competence mainly comes from implicit acquisition processes rather than from explicit grammar teaching.

Besides, nearly all studies focused on the impact of different CF techniques in an ESL context (Bitchener et al., 2005; Sheen, 2007), and as Ellis et al. put it: “There is a clear need for further research, especially in an EFL context” (2008, P. 355). It is hoped that the findings of this study can serve as a source of using effective grammar correction techniques to facilitate EFL students’ learning of “simple past tense”.

**Research questions**
1. Is WCF effective to improve EFL learners writing accuracy?
2. Is there any significant difference in the immediate post-test and delayed post-test performance of red pen group?
3. Is there any significant difference in the immediate post-test and delayed post-test performance of indirect group?
4. Is there any significant difference between the impact of indirect and direct CF techniques on intermediate EFL learners’ writing accuracy comparing their immediate post-tests? If yes, which technique has greater impact?
5. Is there any significant difference between the impact of indirect and direct CF techniques on low-intermediate EFL learners’ writing accuracy comparing their delayed post-test? If yes, which technique has greater impact?

**Research Hypothesis**
1. WCF is not effective to improve EFL learners writing accuracy.
2. There is no significant difference in the immediate post-test and delayed post-test performance of red pen group.
3. There is no significant difference in the immediate post-test and delayed post-test performance of indirect group.
4. There is no significant difference between the impact of indirect and direct CF techniques on intermediate EFL learners’ writing accuracy comparing their immediate post-tests.
5. There is no significant difference between the impact of indirect and direct CF techniques on low-intermediate EFL learners’ writing accuracy comparing their delayed post-test.

**II. Method**

**Participants**

90 EFL learners participated in the study. They were randomly selected from among 135 students taking part in intermediate courses at Iran-Australia Language School in Tehran. The ratio of male to female participants was nearly equal, i.e. 49% males and 51% females to avoid bias caused by sex difference. The sample consisted of a homogeneous group in terms of age, first language, and the English language background. The average age of the participants was 22.7. Their first language was Persian, and their English language proficiency was nearly equal.

**Instruments**

Two tests were employed in the present study. The first test was the Cambridge’s Preliminary English Test (PET) used to ascertain the homogeneity of the participants with regard to their English proficiency. The second was a writing test package which included a pre-test, an immediate post-test, and a delayed post-test in order to measure the participants’ achievement.

**Reliability of the instruments**

The participants’ pieces of writing were evaluated and scored by two raters for assessing inter-rater reliability. To estimate the inter-rater reliability of the test, we calculated the correlation coefficient between the two raters. Table 1 depicts the resulting inter-rater reliability indices.

<table>
<thead>
<tr>
<th>TABLE 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>INTER-RATER RELIABILITY INDICES</td>
</tr>
<tr>
<td>CM test R1</td>
</tr>
<tr>
<td>CM test R2</td>
</tr>
</tbody>
</table>

Note. R1 = first rater; R2 = second rater. ** p < 0.1.
The test is shown to have very high reliability, 0.957, which is statistically significant at $p < 0.1$ level of significance. The content validity of the instruments was assured by a panel of experts.

**Procedure**

Class sessions were held three times a week for 12 weeks in summer 2011, with each session taking 90 minutes. The course was incorporated into a competency-based syllabus, the objective of which was to promote the learner's communicative skills (reading, writing, speaking, and listening). To complete this course of the study, we followed three stages:

**Stage one:** 135 students form 14 intact classes of “Intermediate-Level” in Iran-Australia Language School were given a PET (Preliminary English Test) in order to select a homogenized sample in terms of language proficiency. PET consisted of 69 English language proficiency questions on the four skills of reading, writing, listening, and speaking. Results of this test provided each of the participants with a standardized score out of 100 to see how they performed. Out of the participants who took the test, the eligible ones (those whose scores ranged from one standard deviation above and below the mean on the test) were selected to take part in the study. They were then randomly separated into three 30-member homogeneous groups. The conditions for all groups were exactly the same, except for the method used for the provision of written feedback.

**Stage two:** The experimental group named A received direct WCF in red pen, and those who received indirect CF technique were considered as the experimental group B. Group C (the control group) received no CF for their writings.

In each session, the teacher in the experimental group carried out a specified process to conduct the treatment. The data was collected every three sessions during the course. On the first day, the participants took the pre-test. On the same day, the teacher asked the participants to write about a certain topic at home and bring it the next session. Next session the teacher took the participants' writing home and provided feedback. In the third session, the teacher handed the participants' writing. After the treatment was done, the teacher asked the participants to rewrite the writing and bring them back the following session. In the next session, after collecting the revised pieces of writing, the teacher gave the participants another topic to write at home for the other next session. The same procedure was kept going for the rest of the term. The teacher gathered 12 writing samples in this way.

As for the control group, the teacher handed the participants' writing and asked them to rewrite the texts and bring them the next session. Next session, after collecting the rewrites, she gave another topic for next session. This procedure took about one minute of the class time each session.

**Stage three:**

A pre-test, an immediate post-test, and a delayed post-test were run. Each participant completed 3 in-class writing tasks as part of the writing assessment staged at weeks 1, 12 and 20. Pre-test and immediate-post-test writing tasks were identical, whereas the delayed post-test writing task was of a slight difference in content – a descriptive writing to prompt the students to use the targeted linguistics forms. As for the pre-test and the immediate post-test, the students were asked to write paragraphs in which they described their activities in the preceding day. In the delayed post-test, the participants were required to write about their last trip. The participants were given equal amount of time (15 minutes) for these writing activities.

Since the researchers aimed to use the same topic of pre-test for the immediate post-test, the participants did not receive any feedback on their pre-test writing and even their pieces of writing were not given back to them. For all groups the immediate post-test completed in the last session and took the same amount of time. The immediate post-test was returned to all participants one week after completion without receiving any correction on this occasion.

The delayed post-test was run in week twenty. The teachers agreed not to attract any attention on the targeted forms during the interim period. The produced pieces were returned to the participants one week later.

### III. DATA ANALYSIS

The design used in the current study was true experimental; as a true experimental design allows, the study included a control group, and used stratified randomization. In addition, it conducted a pre-test as well as immediate and delayed post-tests. In order to examine the performance of the three groups in the immediate post-test of writing accuracy a one-way analysis of variance (ANOVA) was conducted. Similarly, to compare the performance of the three groups in the delayed post-test, we run a second ANOVA. Furthermore, to examine the performance of the two experimental groups (group A and group B) on three different tests (pre-test, immediate post-test, and delayed post-test) given to them during the study, the researchers made repeated comparisons using the matched t-test for each group and then corrected the significance level through the Bonferroni test.

### IV. RESULTS

**Performance of the Three Groups in the Immediate Post-Test**

In order to examine the performance of the three groups in the immediate post-test of writing accuracy a one-way ANOVA was conducted. The results are presented in the table 2.
As the table shows, the differences between the groups are significant (Sig=.000). Therefore, the participants in the three groups differed in their performance in the immediate post-test. Post hoc analysis was conducted through the Scheffe test to specify exactly which group was different from others. This is presented in the table 3.

### Table 3

<table>
<thead>
<tr>
<th>(I) groups</th>
<th>(J) groups</th>
<th>Mean Difference (I-J)</th>
<th>SE</th>
<th>Sig.</th>
<th>Lower Bound</th>
<th>Upper Bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scheffe</td>
<td>direct</td>
<td>-1.3333</td>
<td>.22845</td>
<td>.844</td>
<td>-.4356</td>
<td>.7023</td>
</tr>
<tr>
<td></td>
<td>indirect</td>
<td>4.00430</td>
<td>.22467</td>
<td>.000</td>
<td>3.4448</td>
<td>4.5638</td>
</tr>
<tr>
<td></td>
<td>control</td>
<td>-.13333</td>
<td>.22845</td>
<td>.844</td>
<td>-.7023</td>
<td>.4356</td>
</tr>
<tr>
<td></td>
<td>indirect</td>
<td>3.87097</td>
<td>.22663</td>
<td>.000</td>
<td>3.3066</td>
<td>4.4354</td>
</tr>
<tr>
<td>control</td>
<td>direct</td>
<td>-4.00430</td>
<td>.22467</td>
<td>.000</td>
<td>-5.5638</td>
<td>-3.4448</td>
</tr>
<tr>
<td></td>
<td>indirect</td>
<td>-3.87097</td>
<td>.22663</td>
<td>.000</td>
<td>-4.4354</td>
<td>-3.3066</td>
</tr>
</tbody>
</table>

* The mean difference is significant at the 0.05 level.

The results show a significant difference among all the paired groups with regard to their performance in the delayed post-test. This amounts to saying that the three groups differ significantly in their performance on the delayed post-test. A look at the mean differences from the table makes it clear that the group B (the indirect feedback group) has
performed significantly better than the group A (direct feedback group) and the group C (the control group) on the delayed post-test. In addition, the group A performed significantly better than the group C in the delayed post-test.

The calculated effect size based on the guidelines provided by Pallant (2005) was 0.8 which is considered to be a large effect size.

\[ Eta\ squared: \frac{669.409}{759.656} = 0.881 \]

**Red Pen Corrective Feedback and Writing Accuracy**

In order to examine the performance of the two experimental groups (group A and group B) on three different tests (pre-test, immediate post-test, and delayed post-test) given to them during the study, repeated comparisons were made using the matched t-test for each group and then corrected the significance level through the Bonferroni test. Based on the Bonferroni test, the level of significance is first decided at 0.05 and then, since there was going to be three comparisons made in this part, the level of significance was set at \( \frac{0.05}{3} = 0.017 \) (Pallant, 2005). The matched t-tests that were run were interpreted according to the new level of significance (\( p < 0.017 \)). The descriptive statistics information for the performance of group A is provided in table 6.

**TABLE 6**

<table>
<thead>
<tr>
<th>Pair</th>
<th>Pre</th>
<th>M</th>
<th>N</th>
<th>SD</th>
<th>Std. Error Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>pre</td>
<td>8.0333</td>
<td>30</td>
<td>.85029</td>
<td>.15524</td>
</tr>
<tr>
<td></td>
<td>immediate</td>
<td>16.1333</td>
<td>30</td>
<td>.81931</td>
<td>.14958</td>
</tr>
<tr>
<td>2</td>
<td>pre</td>
<td>8.0333</td>
<td>30</td>
<td>.85029</td>
<td>.15524</td>
</tr>
<tr>
<td></td>
<td>delayed</td>
<td>14.0000</td>
<td>30</td>
<td>.83045</td>
<td>.15162</td>
</tr>
<tr>
<td>3</td>
<td>immediate</td>
<td>16.1333</td>
<td>30</td>
<td>.81931</td>
<td>.14958</td>
</tr>
<tr>
<td></td>
<td>delayed</td>
<td>14.0000</td>
<td>30</td>
<td>.83045</td>
<td>.15162</td>
</tr>
</tbody>
</table>

As the table shows, this group has the highest performance in the immediate post-test (\( M=16.13 \)) followed by delayed post-test (\( M=14 \)) and the lowest performance in the pre-test (\( M=8.03 \)).

For the sake of learning about the difference of the performance of group A, a matched t-test was applied to compare the mean scores of the participants in the first group concerning their performance on the pre-test and the two post-tests of writing accuracy. The analysis was conducted to see whether or not there was any significant difference in the performance of the first group in the three above mentioned tests. The results are presented in the table 7.

**TABLE 7**

<table>
<thead>
<tr>
<th>Paired Differences</th>
<th>Mean</th>
<th>SD</th>
<th>Std. Error Mean</th>
<th>Lower</th>
<th>Upper</th>
<th>( t )</th>
<th>( df ) (2-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-immediate</td>
<td>-8.1000</td>
<td>1.26899</td>
<td>.23169</td>
<td>-8.57385</td>
<td>-7.62615</td>
<td>-34.961</td>
<td>29 .000</td>
</tr>
<tr>
<td>Pre- delayed</td>
<td>-5.96667</td>
<td>1.03335</td>
<td>.18866</td>
<td>-6.35253</td>
<td>-5.58081</td>
<td>-31.626</td>
<td>29 .000</td>
</tr>
<tr>
<td>Immediate- delayed</td>
<td>2.13333</td>
<td>1.16658</td>
<td>.21299</td>
<td>1.69772</td>
<td>2.56894</td>
<td>10.016</td>
<td>29 .000</td>
</tr>
</tbody>
</table>

As the table shows, the observed mean differences are significant at \( p=.017 \). This means that participants in the first group had significantly different performances on the pre-test and the two post-tests. Comparing the mean of the pre-test and post-tests makes it clear that the participants performed significantly better in the immediate post-test than the delayed post-test and obviously the pre-test.

**Indirect Corrective Feedback and Writing Accuracy**

The results of the descriptive statistics of the participants in the second group (group B) revealed that the mean scores of the participants in the second group on the pre-test and two post-tests of writing accuracy were 8.06 and 15.96 and 15.60 respectively. These are presented below in table 8.

**TABLE 8**

<table>
<thead>
<tr>
<th>Paired Samples Statistics of Group B</th>
<th>M</th>
<th>N</th>
<th>SD</th>
<th>Std. Error Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pair 1</td>
<td>8.0667</td>
<td>30</td>
<td>.69149</td>
<td>.12625</td>
</tr>
<tr>
<td>pre</td>
<td>15.9667</td>
<td>30</td>
<td>.80872</td>
<td>.14765</td>
</tr>
<tr>
<td>immediate</td>
<td>15.6000</td>
<td>30</td>
<td>.77013</td>
<td>.14061</td>
</tr>
<tr>
<td>Pair 2</td>
<td>8.0667</td>
<td>30</td>
<td>.69149</td>
<td>.12625</td>
</tr>
<tr>
<td>pre</td>
<td>15.9667</td>
<td>30</td>
<td>.80872</td>
<td>.14765</td>
</tr>
<tr>
<td>delayed</td>
<td>15.6000</td>
<td>30</td>
<td>.77013</td>
<td>.14061</td>
</tr>
</tbody>
</table>

© 2013 ACADEMY PUBLISHER
In order to determine the difference of the performance of the participants in the second group (group B) on the pre-test and the two post-tests, we ran another matched t-test to compare the mean scores of the participants in the three tests. The results of the t-test are shown in Table 9.

<table>
<thead>
<tr>
<th>Paired Differences</th>
<th>M</th>
<th>SD</th>
<th>Std. Error Mean</th>
<th>95% Confidence Interval of the Difference</th>
<th>t</th>
<th>df</th>
<th>Sig. (2-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pair 1 pre-immediate</td>
<td>-.790000</td>
<td>1.02889</td>
<td>1.8785</td>
<td>-8.28419 to -7.51581</td>
<td>-.42055</td>
<td>29</td>
<td>.000</td>
</tr>
<tr>
<td>Pair 2 pre-delayed</td>
<td>-.353333</td>
<td>1.13664</td>
<td>2.0752</td>
<td>-7.95776 to -7.10890</td>
<td>-.36301</td>
<td>29</td>
<td>.094</td>
</tr>
<tr>
<td>Pair 3 Immediate-delayed</td>
<td>.36667</td>
<td>1.15917</td>
<td>.21163</td>
<td>-.04318 to .79951</td>
<td>1.733</td>
<td>29</td>
<td>.094</td>
</tr>
</tbody>
</table>

The results of the matched t-test, as shown in Table 9, revealed that the t-observed values is higher than the critical value in two pairs (pre-test vs. immediate post-test and pre-test vs. delayed post-test) indicating that the difference between the performance of the participants in the pre-test and the two post-test was statistically significant. This suggests that the participants in the second group, too, benefited from WCF provided. However no significant difference was found between the performance of the participants of group B on the immediate and delayed post-tests. This amounts to saying that participants in the second group were able to do equally well in the delayed post-test.

V. Discussion

As the results of the current research show, there seems to be a strong bond between providing language learners with error feedback and their writing accuracy. There are several explanations to support such a theory: (1) In line with Schmidt’s (1990) noticing hypothesis only items which are noticed by the learners will be likely to be acquired. Thus, error feedback (whether direct or indirect) will push the learners towards noticing the linguistic problems they are struggling with and that sometime they take for granted. In other words, providing CF will prompt the learners to try and modify their developing interlanguage system in line with the feedbacks provided. (2) Assimilation theory suggests an effective way of learning which occurs by relating new concepts and propositions to existing concepts and propositional framework held by the learner. The learner’s concept and propositional framework is called the individual’s cognitive structure by Ausubel (1986). Providing CF can be viewed as a fruitful strategy in catalyzing the process of assimilation. (3) Working as a kind of scaffold which helps the knowledge to be organized, structured and modified, CF acts as new knowledge to be accommodated into the existing knowledge and prompts the learners to stick the learnt stuff in their long term memory. (4) According to Sweller (1988), cognitive load theory states that the working memory should have as less load as possible in order to optimize learning which occurs in humans and expedite the alternation in long-term memory ideally. Sweller believes that for learning to take place there should be a link between schematic structures of long term memory and new data and if the connection is not made, the learning won’t be lasting and learners will most likely forget the material. Therefore, CF can be beneficial in the sense that it draws learners’ attention specifically to the areas they have difficulty with while freeing their minds to process language content.

An aim of the current study was to compare and contrast two types of WCF namely the direct red pen correction versus the indirect feedback. According to the results, the indirect feedback group acted significantly better than the other two groups (the red pen feedback group and the control group) on the delayed post-test suggesting the lasting effectiveness of the indirect WCF over direct red pen feedback. This confirms the findings of Sheppard (1992), Frantzen (1995), Fazio (2001), and Chandler (2003) who pinpointed CF as a way of improving the accuracy of L2 students’ writing. However, the present findings run counter to that of Truscott (1999, 2007) who claimed that giving feedback has probable repercussions on learners’ capability to write correctly and if beneficial, it is insignificant.

Furthermore, the results of the current study are in harmony with several other studies which claim that the indirect error correction causes either more or equal levels of accuracy in the long run (Ferris & Helt, 2000; Frantzen, 1995; Lalande, 1982; Lee, 1997; Robb et al., 1986). The fact that the indirect CF group was able to do superior in the delayed post-test may imply the superiority of the indirect method of error correction over the course of time. The important issue to be considered in improving learners’ writing accuracy is whether indirect feedback is more promising than direct feedback. From pedagogical point of view, this is an important issue because coding error types may be slower for teachers than just underlining and correcting. There is no doubt that it is relatively painless to just underline or circle errors. Hence the direct method, though less fruitful over time, may be a more handy option for teachers. On the other hand, applying indirect methods of error correction will necessarily call for sufficient linguistic knowledge possessed by students to self-correct errors and also getting used to self edit their own texts. Therefore, using indirect feedback
strategies which, according to the findings of the present study, has a more lasting effect may be suggested for the later stages of learning (probably intermediate and above intermediate levels).

REFERENCES


**Ataollah Maleki** is currently Professor of TEFL at Zanjan Medical Sciences, Iran. His main areas of research include Applied Linguistics, Phonetics, ESP, and Medical Education. He has published extensively with many international journals.

**Elham Eslami** is currently an EFL instructor at Tehran Institute of Technology, Iran. Her main area of research is TEFL.