

Learning Corrected Target Forms: Students' Corrective Feedback through Teacher's Written Corrective Feedback in Iranian EFL Context

Fatemeh Esfandiar

Islamic Azad University, Ayatollah Amoli, Science and Research Branch, Amol, Mazandran, Iran

Baqure Yaqubi

University of Mazandaran, Babolsar, Mazandran, Iran

Amir Marzban

Islamic Azad University, Qaemshahr Branch, Qaemshahr, Mazandran, Iran

Abstract—As a result of Truscott's (1996) claim that written corrective feedback (WCF) should be abandoned because it is both ineffective and harmful, different opinions about the value of WCF have been voiced. The aim of this article was to investigate (1) the possible effect of WCF on accurate use of quantifiers over a 2-month period; (2) the possible differential effectiveness for three types of feedback. Seventy EFL students at junior high school in Iran formed a control group and three experimental groups (1) those who received direct WCF with written metalinguistic explanation; (2) direct corrective feedback only; (3) indirect WCF. The subject in experimental groups accomplished a collaborative task (correcting five weak examinees' papers) in pairs. Then on three occasions (pre-test, immediate post-test and delayed post-test) the participants were asked to complete seventeen grammar items and write sentences about the given picture. The study found that each of the WCF groups significantly outperformed the control group on immediate post-test and this level of accuracy was retained 2 month later. Furthermore, no meaningful difference in effectiveness was found between the three WCF groups.

Index Terms—written corrective feedback, direct and indirect feedback, metalinguistic feedback

I. INTRODUCTION

In 1996, Truscott claimed that WCF should be abandoned because it is ineffective and harmful and then, in 1999 he admitted that further research should investigate which method techniques or approaches to error correction may have value. His suggestion was in response to Ferris (1999) who argued that Truscott's claims were premature because the evidence he proposed had methodological problem both in design and analysis. Then she suggested that further research is required to investigate the efficacy of WCF. Despite the call for empirical evidence on the efficacy of WCF by both Truscott (1996, 1999, 2007) and Ferris (1999), a number of researchers who believe in effectiveness of WCF in helping language learners improve the accuracy in the use of linguistic features in which errors frequently occur, have proceeded to discover the relative efficacy of different types of direct and (or) indirect WCF (e.g. Bitchener, 2008; Bitchener and Knoch, 2008, 2010a, 2010b; Sheen, 2007; Van Beuningen et al., 2008, 2012).

One problem that draws the interest of the researchers is the need for examining the effect of different types of feedback on certain grammatical forms, Truscott (1996, 1999), Ferris (2004) Bitchener et al. (2005), Bitchener (2008), Bitchener & Knoch ((2008, 2009) point to the fact that different linguistic forms might be acquired differently, and a particular type of CF might be effective for some structures but not for some other forms in another word, different linguistic categories may not be treated as the same; for example, Bitchener et al. (2005) found that the two types of direct WCF used in their study did not have impact on accurate use of preposition but the same types of WCF enabled learners to use past simple tense and the definite article with meaningfully great accuracy and they hypothesized that direct WCF might be useful for treating just some errors. As Bitchener et al.(2005) and Bitchener (2009) suggested further research is required to investigate the impact of WCF on other linguistic error categories. Therefore, the first aim of the present study was to discover the role of WCF in treating errors in the use of quantifiers: the countable quantifiers such as a few, many, and some or a lot of used before plural countable nouns and the uncountable quantifiers utilized before uncountable (singular) nouns.

Both Ferris (1999) and Truscott (1999) agreed that in future research, attention should be focused on investigating the long-term improvement of WCF. This call has resulted in a an ever-expanding number of studies attempting to explore the long-term benefits of WCF (e.g. Sheen, 2007; Bitchener, 2008; Bitchener& Knoch, 2008, 2009, 2010a, 2010b; Van Beuningen et al., 2008, 2012). Among these studies, Van Beuningen et al. (2008, 2012) and Bitchener and Knoch (2010b) reported that, while positive short effects were found for both direct and indirect corrective feedback,

only direct CF option had a significant long-term effect. According to Bitchener (2012), it would be premature to assume this position that the evidence from the recent studies is sufficient to claim in favor of direct CF. Addressing to this issue that the results from these studies have varied, further research is required before making any robust conclusions. Therefore, the second aim of this study was to see if possible effect on WCF options is retained 2 month later.

Another issue to be considered is the impact of two CF methodologies involved direct and indirect CF. A range of claims in support of direct and indirect CF have been made for many years. Some researchers such as Ferris (2006) and Lalande (1982) have reported a benefit for indirect CF. Some others such as Chandler (2003) Van Beuningen et al. (2008), and Bitchener & Knoch (2010) have claimed that in their comparison they found direct CF performed better over indirect CF. On the other hand, some studies such as Frantzen (1995) and Robb et al. (1986) have found no significant difference between two methodologies. These contradictory findings push the researchers of the present study to investigate the effectiveness of direct and indirect CF. Therefore, the third aim of the present study was to investigate whether particular types of WCF facilitate greater accuracy.

In a quasi-experimental design with two occasions (pre-test, immediate post-test) the researchers of the present study measure the effectiveness of three types of WCF and to measure retention over time a delayed post-test was incorporated into design of this study. The study was conducted with 72 junior high school EFL students (Iran, Amol). Assigned to three experimental groups and one control group (direct WCF with written metalinguistic explanation; direct corrective feedback only; indirect WCF; no corrective feedback), the subjects in experimental groups accomplished a collaborative task (correcting five weak examinees' papers) in pairs. Then on three occasions (pre-test, immediate post-test and delayed post-test) the participants were asked to complete seventeen grammar items and write some sentences about the given pictures. Participants in control group did not complete the task so they received no WCF in this study. Before presenting the findings of this study, we will first examine theoretical perspectives for and against WCF then we will offer theoretical objections to effectiveness of WCF raised by Truscott and finally, we aim to review the empirical studies on (in) effectiveness of WCF and the relative efficacy of different types of WCF options.

II. LITERATURE REVIEW

A. *Empirical Studies on the Effectiveness of Written CF*

Early studies divided in to two categories. A set of studies (e.g. Ferris, 1997; Ashwell, 2000; Ferris and Roberts, 2001) worked on the role of CF in revision studies in which researchers compared two versions of the same text. Revision studies have shown the accuracy improvement in a particular piece of writing. However Truscott and Hsu (2008) argued that the finding of these studies cannot be considered as evidence of learning because in their study, Truscott and Hsu (2008) found the group who received WCF for revision significantly outperformed over the content group one week later. Then they asked both group to write a new writing, and result demonstrate that both groups were the same and they concluded that correction did not have any effect on writing development.

The other set of studies (e.g. Polio et al., 1998; Sheppard, 1992; Chandler, 2003; Semke, 1984) were conducted to determined learning effect of CF in new pieces of writing. Some of early studies have resulted in different findings for example while Kepner (1991), Polio et al. (1998), Semke (1984), and Sheppard (1992) found no effectiveness for CF, Chandler (2003) found direct WCF has the impact on the development of students' accuracy; Lalande (1982) reported advantage for indirect CF but this difference was not statistically significant; and Rob et al. (1986) revealed advantage for all four CF groups but they found no significant difference between them.

B. *Studies on Relative Effectiveness of Direct and Indirect CF*

Lalande (1982) studies 60 intermediate German FL learner at a USA university over ten weeks. She compared one group that received indirect coded CF to another that received direct CF. The result found an advantage for indirect CF over direct error correction. Therefore, she concluded that indirect group was engaged in more form focused activities than direct group. However, the reported difference in accuracy development was not statistically significant. On the other hand, Chandler (2003) who divided 36 music students into four groups (direct CF; indirect underlining +marginal CF; indirect marginal CF; indirect underlining CF) found that although there was no statistically significant difference between direct CF and underlining CF groups, direct CF was the best approach for improving the students' accuracy in writing. She also concluded that because students saw immediately the correction of their error, they would be able to "internalize the corrected form better" (p.291). Frantzen (1995) also carried out a 15-week study with 67 intermediate Spanish learners at a U.S. university. She used an indirect CF group and a direct group that also had extensive grammar reviews. She found no significant difference between the groups on accuracy in their writing. She then concluded that indirect approach is sufficient for accuracy development, at least in this case. All of these studies had serious problem in their design and analysis (see Ferris 2004 and Van Beuningen et al. for a review of these methodological issues). Van Beuningen et al. (2008, 2012) and Bitchener and Knoch (2010b) however in their recent studies claimed that they tried to avoid the design and execution shortcoming because they included a control group to the design of their studies. They also investigated the learning potential of WCF in new pieces of writing.

Van Beuningen et al. (2008) carried out a 3-week study. 62 Dutch multilingual secondary school learners. The participants involved: 1) direct CF; 2) indirect CF; 3) writing practice (control group1); self-correction revision (control

group 2). Van Beuningen et al.'s (2012) 6-week study was conducted in the same condition as Van Beuningen (2008) study. However in this study they used 268 participants. Bitchener and Knoch 2010b also did a 10-week study with 63 advanced learners at a large university in the USA. All these studies found that whereas direct and indirect CF proved to have equal short-term effect in developing learners' accuracy, only direct WCF had a more significant long-term effect than indirect WCF.

C. Studies Comparing Direct CF Types

Relative effectiveness of different types of direct CF drew the interest of some researchers (e.g. Bitchener et al. 2005; Bitchener 2008; Bitchener & Knoch, 2008, 2010a; Sheen, 2007). Bitchener et al. 2005 investigated whether two types of direct WCF (direct CF; direct CF+ 5 minutes teacher-student conference) on three types of errors led to increased development in new pieces of writing. The study found a significant effect for the group who received direct WCF plus oral metalinguistic explanation for simple past and definite article but no effect on accuracy improvement for preposition. They concluded that including metalinguistic explanation "would help learners notice the difference between their errors and the correction they receive" (p. 201), and this noticing as Schmidt (1990) stated, is a crucial factor in long-term acquisition. Since Bitchener et al. (2005) did not consider the effectiveness of WCF in targeting different functions of articles, Bitchener (2008) compared the effect of the different types of direct CF for two functional uses of English article systems. He used three types of direct CF (direct CF+ written and oral metalinguistic explanation; direct CF+ written metalinguistic CF; direct CF only) as well as a control group. They found that all CF groups revealed improved accuracy in immediate post-test but participants in written and oral metalinguistic CF group and direct CF group could retain this level of accuracy two months later. Bitchener and Knoch (2008) they repeated Bitchener's (2008) study by including an additional 69 learners. When the study was extended, no differences were reported between the same CF groups and all three CF groups were significantly different from the control groups. According to Bitchener and Knoch (2010a), it might be possible that "the larger sample size eliminated the difference in effect between group 2 and other two treatment groups in the first study by Bitchener (2008)" (p. 199).

Bitchener and Knoch (2009) also conducted a six-month study with three different types of direct CF (direct CF+ written and oral metalinguistic explanation; direct CF+ written metalinguistic CF; direct CF). They found that all three groups significantly improved their accuracy after treatment and suggested that for learners at low intermediate proficiency level, providing a single CF alone might be sufficient. Bitchener and Knoch (2010a) carried out a more longitudinal study in which they aimed to investigate the effect of three direct CF types (direct CF+ written and oral metalinguistic explanation; direct CF+ written metalinguistic CF; direct CF). The study found that all three CF groups outperformed the control group on all post-tests. On the other hand, Sheen (2007) who did a 2-month study on the relative effectiveness of two types of direct CF (direct CF only; and direct CF+ metalinguistic explanation) found that both CF groups performed better than the control group only on immediate post-test but in delayed post-test, participants of direct CF plus metalinguistic explanation group outperformed those in direct CF and control groups. Then she concluded that direct CF with metalinguistic explanation was superior to direct CF alone.

III. THE STUDY

The study was designed to investigate the possible effect of WCF on the accurate use of quantifiers over a two month period. Thus, a pre-test post-test design was used (a pre-test at the beginning of the two month period; and post-tests after one week and two months). The second aim of this study was to discover if there is a possible differential effectiveness for three types of written feedback. Two research questions were therefore formulated to investigate these aims.

1. Does accuracy in the use of quantifiers vary according to three types of WCF?
2. Does accuracy in the use of quantifiers vary over time according to three types of WCF?

IV. METHODOLOGY

A. Participants

Based on the consensus among researchers regarding the larger the size of the sample, the greater its precision, the researcher invited a total 110 students from four intact classes in three junior high schools in Amol, Iran, to participate in this study at the first stage. All the participants were studying in grade three. There were 50 males and 55 females students in the sample. In order to make the sample more homogeneous, the researcher used the pre-test scores and this homogeneity was not statistically meaningful. Based on the result of the pretests (grammar test& picture description test) in each class, subjects who placed between ± 1 standard deviation above and below the mean were selected as the main subjects of this study and 40 students were excluded from the study because they had extremely high, or extremely low scores on the test. Then the participants were divided into 3 experimental groups and 1 control group.

B. Target Structures

Compared with earlier studies on the value of written WCF (see Ferris, 2006), that fifteen linguistic error categories have been investigated, this study examined the effect of targeting the use of quantifiers. This structure was targeted

because ten years' teaching experience in grade three of junior high school has shown that students experience difficulties in the use of quantifiers. Difficulties occur when deciding whether countable quantifiers (quantifiers used for countable plural nouns, e.g. many, a few) or non-countable quantifiers (quantifiers used for non-countable nouns, e.g. much, a little) should be used; whether singular nouns or plural nouns should be used after quantifiers like some or a lot of; the students may have difficulty deciding whether singular or plural form of to be verb (is, are) should be used for a particular quantifiers. Accuracy in the use of this structure in the grammar pre-test (13.83) and picture description pre-test (10.93) revealed that students have only partial mastery of the use of the quantifiers.

C. Treatment

One week after pre-test session, the treatment session took place. In this session, the three WCF groups received one type of the WCF on errors made in accurate use of English quantifiers and then in group of two (during student-student interaction) gave direct and metalinguistic WCF on their peers' papers. Participants in WCF group 1, received the corrected forms of the target structure through a key answer sheet. Students were asked to compare their peers' responses in the exam papers with the teacher's key answer sheet and find all existing linguistic errors, then, they were told to give direct corrective feedback with metalinguistic explanation to their peers' papers, as illustrated in Example 1:

Example (1):

The original text: *How many uncle does he have?

Teacher's direct WCF: How many uncle does he have?

Participant's CF: how many uncles because after how many we use plural noun.

In WCF group 2, in addition to teacher's direct CF (key answer sheet), participants were given teacher's written metalinguistic explanation in which the teacher explained the errors made in accurate use of quantifiers. Example (2) illustrated teacher's meta-linguistic CF.

Example (2): metalinguistic CF

Q1. Because bread is uncountable, you should use a little.

Q2. Because book is a countable noun and you should use How many for making question.

In WCF group 3, the researcher underlined all linguistic errors targeted on quantifiers but the correct forms were not provided by the teacher. Instead, pairs were left to work out and correct the erroneous features themselves. For this purpose they were allowed to search the answer from external linguistic resource such as their pair, grammar book or textbook and give their peer's paper metalinguistic CF. It should be mentioned that in all experimental groups, each student must correct the peers' paper in collaboration with another member and their pair-talk will be recorded for analysis. The researcher of the present investigation did not consider any time limit for collaborative peer feedback in treatment session because collaboratively correcting the peers' paper may vary from pair to pair, so it might take about 45 to 90 minutes.

D. Instruments

All three grammar tests (grammar pre-test, grammar post-test 1 and grammar post-test 2) required participants to answer some items involve: 7 multiple choice items; 5 filling in the blank items; 5 making and answering question items that all are taken from a popular work book named 'Tajik'. These 17 grammar items were the same in pre-test, post-test 1 and post-test 2. Twenty minutes was given for completing 17 items in each test, i.e. pre-test, grammar post-test 1 and grammar post-test 2. Besides, the last item of these three tests was a writing activity and required the students to describe what they saw in the given pictures. In the pretest and immediate posttest, participants were given extra five minutes to write five sentences about the pictures, and in delayed posttest, extra ten minutes was given for writing ten sentences about the pictures. Since the vocabulary items that participants needed to use for writing activity were familiar for them, they were told that they could not use a dictionary or have the teacher aid during the activity. The given illustrations in each of the test were different. The picture descriptions were selected because the range of people and objects illustrated had the potential to provide obligatory opportunities for students to use different forms of English quantifier. In order to record the conversation between the participants in experimental groups an MP3 player was utilized. A Nelson Test was also used in pilot study to determine the reliability and validity of the teacher made test.

E. Procedure

As an initial data collection step, the researcher of this study sought permission from the authorities of the junior high schools in Amol and finally, among ten schools that the researcher tried to get permission, the management of two male and one female guidance schools allowed the researcher to utilize their students for carrying out the study. Because the two schools that participated in this study were male schools, the researcher needed the aid of two male teachers; therefore the next step was to obtain consent from two male teachers; and fortunately among five male teachers, two of them accepted to cooperate with the researcher. After getting the consent of the authorities and the two teachers, the pretest was piloted by 21 students who were randomly selected among the subjects in the three schools. During this pilot phase, test sections were revised and prepared for the main subjects.

The present study was mainly concerning with investigating and analyzing whether using collaborative peer feedback through the teacher's direct and indirect WCF on students' exam papers may lead to learning corrected grammatical forms. For this purpose, based on the pre-test scores, the subjects were screened into four groups involved direct CF

group; direct CF with metalinguistic CF group; indirect CF group and no CF group. Then the subjects in CF groups completed an editing task involved participants' classmates' exam papers. In fact, instead of using a standard text with errors implanted as Lee (1997) suggested, the researcher used the students' papers to examine the participants' performance in error correction of a their classmates' exam papers where errors occur naturally. Participants were asked to provide their peers' papers with direct and metalinguistic feedback. Subjects in control group completed the tests (i.e., pre-test and post-tests) only. They did not perform the editing task (i.e., correcting their peers' papers) and did not receive any feedback but instead followed normal classes. Students in CF groups received a digital voice recorder to be placed between each pair. Then after one week all groups were given the post-test1 in order to recognize the subjects' ability in accurate use of quantifiers after treatment. Finally, after two months, the second post-test were given to four groups in order to probe accuracy in the use of quantifiers.

A few days prior to the pretest, the students of all groups were provided with information about the study. Each experimental group (in separate sessions) was explained the type of WCF that they were asked to give their peers' papers and they were emphasized that this activity would be done in pair. When the researcher made sure that the students became familiar with the approach used in the study, the two teachers who took part in this study were also briefed on the procedures. This study utilized quantitative methods of data analysis with the aim of investigating whether using collaborative peer feedback through teacher's direct and indirect WCF on students' exam papers can internalize grammatical forms. In this study corrective feedback was the independent variable and three sets of tests (pre-test, post-test1 and post-test2) were the three dependent variables. The data collected via the pretest, posttest and delayed posttest was first coded and then SPSS (Statistical Package for the Social Sciences) datasets were used for descriptive and inferential statistics. Descriptive statistics for the pretest and the two posttests were calculated separately for the four groups. To establish whether the differences in the four groups' scores on the pretests were statically significant, a one way ANOVA was performed. Because, there were no statistically significant differences on the pretest scores, one- way ANOVA was chosen again to address the research question. Post-hoc multiple comparison tests were computed to isolate the exact points where the significant differences lay among the group.

V. RESULTS

Table 1 below shows the descriptive statistics for the three treatment groups and the control group at the three different testing times (pre-test, immediate post-test and delayed post-test).

TABLE 1
DESCRIPTIVE STATISTICS FOR MEAN TEST SCORES BY GROUP AND TESTING PERIOD

		Direct CF (N= 16)		Meta CF (N=16)		Indirect CF (N=16)		Control (N=22)	
		M	SD	M	SD	M	SD	M	SD
Pre-test	Grammar	13.68	0.91	13.88	1.09	13.79	1.16	13.93	1.03
	Picture description	10.00	4.69	11.27	3.64	11.18	2.00	11.13	2.91
Post-test 1	Grammar	18.32	1.93	19.54	0.72	19.50	1.23	13.23	1.81
	Picture description	18.12	3.22	19.11	1.84	19.81	0.54	10.22	4.62
Post-test2	Grammar	18.14	2.23	19.44	0.97	19.39	0.75	13.34	1.23
	Picture description	17.90	3.71	19.05	1.31	19.12	1.13	10.95	4.01

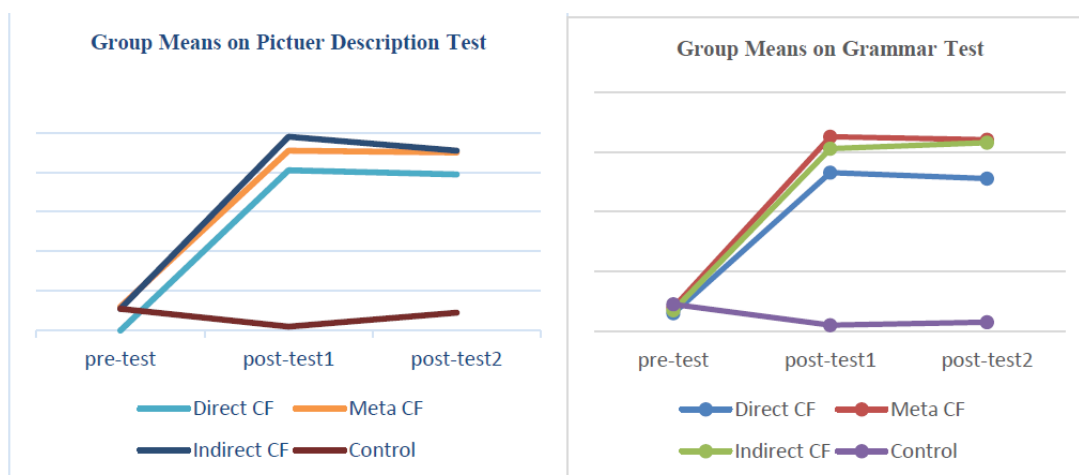


Figure1: Accuracy of whole sample across the three grammar tests and picture description tests

As the result of the grammar pre-test was revealed in table1, the means of all groups in grammar were so close to each other ranging between 13.68 to 13.93, i.e. there is not a large numerical difference among the measures of four

groups in grammar pre-test. The same story can be seen about picture description pre-test that is the means of all groups were so close each other ranging between 10 to 11.27, i.e. there was not a large numerical difference among the means of the four groups in picture description pre-test.

Table 1 also reveals the means of all four groups in grammar post-test1. The means for all experimental groups were so close to each other ranging from 18.32 to 19.54 while the mean of control group (13.23) is numerically so different from the experimental groups. Similarly in picture descriptive post-test 1, as the result of table 1 suggests, the means of all experimental groups are so close to each other ranging from 18.12 to 19.81 while the mean of control group (10.33) is numerically so different from the experimental groups. According to Table 1, the means of all experimental groups in both grammar and picture description post-test 2 were so close to each other ranging from 18.14 to 19.44 in grammar and 17.90 to 19.12 in picture description while the mean of control group both in grammar (13.34) and picture description (10.95) post-test2 was numerically so different from the experimental groups. In order to compare the treatment groups' test scores, a series of one-way ANOVA was computed. Table 2 below shows.

TABLE 2
ONE-WAY ANOVA FOR GRAMMAR AND PICTURE DESCRIPTION PRE-TEST AND POST-TESTS
BETWEEN GROUPS

		df	F	sig
Pre-test	Grammar	3	0.188	0.904
	Picture description	3	0.514	0.674
Post-test 1	Grammar	3	78.716	0.000
	Picture description	3	40.871	0.000
Post-test 2	Grammar	3	87.90	0.000
	Picture description	3	35.63	0.000

As Table 2 shows, regarding the grammar ($F = 0.188$; $df = 3$; $p = 0.904$) and picture description pre-test ($F = 0.514$; $df = 3$; $p = 0.674$), the ANOVA did not reveal a significant difference across the four groups because the calculated level of significance was more than 0.05. In term of the grammar post-test 1, as the ANOVA results suggested ($F = 78.716$; $df = 3$; $*p = 0.000 < 0.05$), there were significant differences among four groups. To statistically examine the exact places of differences among groups, post hoc multiple comparison tests were performed. These indicated that probability (sig) among experimental groups was much more than 0.05, i.e. there was no significant and meaningful difference among experimental groups whilst the probability (sig) among control group and experimental groups is 0.000 i.e., there was a significant and meaningful difference among control group and experimental groups.

Regarding the picture description post-test1, according to Table 2, the ANOVA test revealed that there was a significant difference among four groups at the level of 0.05 ($F = 40.871$; $df = 3$; $*p = 0.000 < 0.05$). Post hoc multiple comparison further showed that the statistically significant differences were located between the control group and three experimental groups because probability (sig) among experimental groups was much more than 0.05, i.e. none of the three experimental groups differed from each other whilst the probability (sig) among control group and experimental groups was .000, i.e. the control group was statistically significant different from all other groups experimental groups.

In term of the grammar post-test 2, as can be seen in Table 4.2, the ANOVA indicated that since the calculated level of significance (0.000) was lower than .05, there was a significant difference among groups ($F = 87.204$; $df = 3$; $*p = 0.000 < 0.05$). In order to locate the exact places of differences in subjects' performance, a post hoc test was run. The result showed that at the time of delayed post-test (post-test 2), participants in the three treatment groups significantly outperformed those in the control group, but that the three treatment groups did not differ from each other.

Similarly, one-way ANOVA showed statistically significant group differences in the picture description post-test2, ($F = 35.631$; $df = 3$; $*p = 0.000 < 0.05$). Post hoc multiple comparison test revealed that probability (sig) among experimental groups was much more than .05, i.e. there was no significant and meaningful difference among experimental groups but the probability (sig) among control group and experimental groups was .000, i.e. there was a significant and meaningful difference among control group and experimental groups.

VI. DISCUSSION

The first research question investigated whether or not there was a differential effect on accuracy for the different types of CF. It found that all three treatment groups (direct CF only; direct CF in conjunction with metalinguistic CF; indirect CF groups) outperformed the control group in the immediate post-test. In other words, all three CF groups gained improved accuracy scores immediately after they had received WCF on their editing task. The results also showed that there was no significant difference between the three treatment groups on the immediate post-test. Therefore, it might be suggested that any one of these three types of WCF could have the same positive effect.

This finding is in line with some other research that investigated the relative short effect of direct and indirect CF. Van Beuningen et al. (2008, 2012), and Bitchener and Knoch (2010b) for instance, found that both direct and indirect CF groups made the accuracy gains turned out to be significant. Moreover, they significantly outperformed the control groups. On the other hand, although Chandler (2003) found no statistically significant difference between direct and indirect CF, claimed that direct approach was the best option for the students' accuracy. Lalande (1982) reported advantage for indirect CF but she did not find statistically significant difference between direct and indirect CF.

According to the results of immediate post-test, the simple provision of error correction was just as effective as the additional provision of metalinguistic explanation. The same result was reported by Bitchener (2008) Bitchener, Knoch (2008, 2009, 2010a), and Sheen (2007). In each of these studies, the effectiveness of direct CF only and metalinguistic explanations were evident in immediate post-test. Bitchener et al. (2005) however, reported different results. They did not find any statistically significant effect for direct CF alone. Bitchener et al. found that direct CF with teacher's metalinguistic comments on errors resulted in significant gains in accuracy in two grammatical forms. They believed that participants who received direct CF alone did not have opportunity to discuss their corrected errors.

It can be contented that the finding of this study in general gain support from previous research in term of learning potential of CF regardless of the type of correction (Lalande, 1982; Rob et al., 1986; Ferris, 1997; Lee, 1997; Ashwell, 2000; Ferris and Robert 2001; Chandler, 2003; Bitchener et al 2005; Bitchener, 2008; Bitchener and Knoch, 2008, 2009, 2010a, 2010b; Sheen, 2007; Ellis et al., 2008; Van Beuningen et al., 2008, 2012). However, some research findings claimed the opposite (Semke, 1984; Kepner, 1991, Sheppard, 1992; Truscott, 1996, 1999; 2007; Truscott & Hsu, 2008). For example, Truscott's (1996) review of early studies by Semke (1984), Kepner (1991), and Sheppard (1992) claimed that CF did not have a significant effect on improving language accuracy. The finding of this study however did not confirm his claim. Truscott (2007) has argued not only that CF is ineffective but also it is harmful to learners' accuracy development because it encourages learners to avoid producing complex structure and leads to simplified writing. The findings of this study oppose this claim. Only in CF groups, the new sentences written by participants in post-tests were more accurate and complex than what they wrote in pre-test. For example, sentences such as "there are three boys in the picture." or "there are breads on the table." Were frequently observed in participants' writing before the treatment session but in post-test sessions, pupils who received CF could improve accurate use of quantifiers and produce more complex writing such as "I see a few birds in the picture." or "there is some cheese on the table."

The second research question was formulated to investigate whether this level of accuracy was retained 2 month later. It revealed that all three CF groups significantly performed better than the control group in the delayed post-test. Results also found no statistically significant difference among three types of CF. in other words, the achieved level of improvement across the 2 months period were retained by all three CF groups. This finding is in contrast with some research (e.g. Bitchener and Knoch, 2010b; Van Beuningen et al. 2008, 2012) which claimed that direct error correction had a more significant long-term effect than indirect CF. However Rob et al. (1986) found no significant difference among direct and indirect CF groups and claimed that all CF groups improved over time.

Regarding the effectiveness of two types of direct CF (direct CF alone, and direct CF plus written metalinguistic CF), the result of this study is in line with Bitchener& knoch (2008, 2009, 2010a) which investigated the relative effectiveness of three types of direct CF and found that all direct CF groups (direct CF alone; direct CF with written and oral metalinguistic CF; and direct CF plus written metalinguistic CF) significantly improved over time. The findings of this study however, contrasted with other research (e.g. Bitchener et al. 2005; Sheen, 2007; Bitchener, 2008). For instance, Bitchener et al. (2005) found a significant effect for the group who received direct CF with metalinguistic comments but no significant effectiveness was reported on accuracy improvement for direct CF alone and control groups. Similarly, Sheen (2007) reported that participants in direct metalinguistic CF group outperformed direct CF and control groups in delayed post-test. Then, she concluded that direct CF only promotes awareness but direct CF with metalinguistic explanation promote both awareness and understanding Bitchener (2008) although, reported that only two of three directCF groups (i.e., direct CF alone; and direct CF with written and oral metalinguistic CF group) performed better than control group but he found no statistically significant difference between those who received direct CF plus written metalinguistic CF and control. Bitchener claimed that "the single provision of written metalinguistic explanation may not have been sufficient for it to have had a significant effect" (p.114).

As it can be seen, reporting different results in effectiveness of CF types was not observed only in our study. These differences in effect were evident in three other studies (Bitchener et al. 2005; Sheen, 2007; Bitchener, 2008) Bitchener and Knoch (2010a) explained reasons such as the effect of sample size, getting additional input between pre-test and post-tests or differences in amount and delivery of CF. It seems that in my study, the frequency and the way of offering and receiving CF in treatment session might be reasons for the differences. In the present study, before immediate post-test, participants in CF groups were allocated 90 minutes to complete an editing task in which they were asked to correct five papers of their classmates' pre-test while receiving one type of the teacher's feedback. But most of studies mentioned above, (e.g.; Van Beuningen et al. 2008, 2012) allocated fewer time for treatment session. For example, Bitchener et al. (2005) Sheen (2007) Bitchener (2008) Bitchener and Knoch (2008, 2009, 2010a, 2010b) asked participants to look at the correction made by the teacher for 5 minutes. Regarding Van Beuningen (2008, 2012) studies, students were allocated 20 minutes to revise all errors corrected by the researchers.

Unlike previous study in which participants were asked to revise only their own pre-test writing, in this study students were asked to work in pairs and correct five pre-test papers completed by their classmates. Additionally, each pair had to give direct oral and written metalinguistic CF to each paper. Thus, they frequently needed to refer to their teacher's WCF. Besides, repeatedly articulating the corrected forms during correction, as Lantolf (2006) mentioned, may facilitate memorization the new forms or consolidate structures that have already been learned. Regarding the importance of repetitions, Storch (2007) also stated that "a repetition by the learner of the corrective feedback offered indicates that the learner has 'noticed' the correction given or the suggestion made" (p.155).

VII. CONCLUSION

The aims of this study were to investigate the efficacy of WCF on targeted linguistic error categories (English quantifiers) over time and to explore if there was a differential effect on accuracy for three corrective feedback types (direct CF; direct CF plus metalinguistic CF; indirect CF). The study found that participants who received WCF performed better than those who received no CF and this level of improvement continued over a 2-month period. With respect to the second aim, the study revealed no difference in effect between the three treatment groups.

The finding of this study is good news for teachers who spend a lot of time and energy providing error feedback on their students' papers, it has shown that a focused approach to the treatment of linguistic errors does not involve extensive amount of class time and may be resulted in helping students be able to make effective use of WCF and reduce the error frequency of the targeted categories. It can be concluded that the present study has benefits for both students and teachers. Regarding students, it may improve students' ability to give and receive feedback about each other's writing and encourage them to take more responsibility for their own learning. Teachers may gain better insight about different types of WCF and pair work. They may also learn some practical advice about using WCF options in their classrooms.

Despite these positive findings, like any kind of research, the present study suffers from a number of limitations which will pose inevitable restrictions upon the generalization of its result. First, because of the difficulty in using subjects over a long time, the sample size, while acceptable, was smaller than one would have wished for. Therefore it lacks generalizability powers. Second, the researchers used students' pre-test scores for homogenizing the participants and clearly it is a restriction upon the generalization of its result. Third, the writing test in picture description part of pre-test and post-tests involved the production of relatively short text (only one sentence for each picture). Fourth, the study examined the effect of CF on just accurate use of English quantifiers and clearly the result cannot be generalized to other areas of grammatical accuracy.

With regard to the present study, the following areas may be worthy of future investigation. (1) This study only considered the effectiveness of WCF on the accurate use of English quantifiers. However, it would be suggested that further research is needed to investigate the extent to which the positive findings of this study apply to other linguistic error categories. (2) This study only investigated the effectiveness of two types of direct CF and one type of indirect CF. Although it has filled another gap in the research, further studies are now required to examine the relative merits of different types of indirect CF and Direct CF alone.

REFERENCES

- [1] Ashwell, T. (2000). Pattern of teacher response to student writing in a multiple-draft composition classroom: Is content feedback followed by form feedback the best method? *Journal of Second Language Writing*, 9, 227-258.
- [2] Bitchener, J. (2008). Evidence in support of written corrective feedback. *Journal of Second Language Writing*, 17, 102-118.
- [3] Bitchener, J. (2012). A reflection on 'the language learning potential' of written CF. *Journal of Second Language Writing*, 21, 348-363.
- [4] Bitchener, J. & Ferris, D. (2012). *Written corrective feedback in second language acquisition and writing*. Routledge. New York.
- [5] Bitchener, J. & Knoch, U. (2008). The value of written corrective feedback for migrant and international students. *Language Teaching Research*, 12, 409-431.
- [6] Bitchener, J. & Knoch, U. (2009). The relative effectiveness of different types of direct written corrective feedback. *System*, 37, 322-329.
- [7] Bitchener, J. & Knoch, U. (2010a). The contribution of written corrective feedback to language development: A ten month investigation. *Applied Linguistics*, 31, 193-214.
- [8] Bitchener, J. & Knoch, U. (2010b). Raising the linguistic accuracy level of advanced L2 writers with written corrective feedback. *Journal of Second Language Writing*, 19, 207-217.
- [9] Bitchener, J., Young, S., & Cameron, D. (2005). The effect of different types of corrective feedback on ESL student writing. *Journal of Second Language Writing*, 14, 227-258.
- [10] Chandler, J. (2003). The efficacy of various kinds of error feedback for improvement in the accuracy and fluency of L2 student writing. *Journal of Second Language Writing*, 12, 267-296.
- [11] Ferris, D. (1997). The influence of teacher commentary on student revision. *TESOL Quarterly*, 31, 315-339.
- [12] Ferris, D. (1999). The case of grammar correction in L2 writing classes: A response to Truscott (1996). *Journal of Second Language Writing*, 8, 1-11.
- [13] Ferris, D. (2004). The "grammar correction" debate in L2 writing: Where are we, and where do we go from here? (and what do we do in the meantime...?). *Journal of Second Language Writing*, 13, 49-62.
- [14] Ferris, D. (2006). Does error feedback help student writer? New evidence on the short – and long term effect of written error correction. In K. Hyland & F. Hyland (Eds.), *Feedback in second language writing: context and issues* (pp. 81-104). Cambridge: Cambridge University Press.
- [15] Ferris, D., & Robert, B. J. (2001). Error feedback in L2 writing classes: How explicit does it need to be? *Journal of Second Language Writing*, 10, 162-184.
- [16] Frantzen, D. (1995). The effect of grammar supplementation on written accuracy in an intermediate Spanish content course. *Modern Language Journal*, 79, 329-344.
- [17] Kepner, C. (1991). An experiment in the relationship of types of written feedback to the development of second language writing skills. *Modern Language Journal*, 75, 305-313.
- [18] Lalande, J. F. (1982). Reducing composition errors: An experiment. *Modern Language Journal*, 66, 140-149.

- [19] Lantolf, J. P. (2006). Sociocultural theory and L2. State of the art. *Studies in Second Language Acquisition*, 28, 67-109.
- [20] Lee, I. (1997). ESL learner's performance in error correction in writing. *System*, 25(4), 465-477. *Language Teaching Research*, 8, 55-81.
- [21] Polio, C., Fleck, C., & Leder, N. (1998). "If I had more time": ESL learners' changes in linguistic accuracy on essay revision. *Journal of Second Language Writing*, 7, 43-68.
- [22] Robb, T., Ross, S., & Shortreed, I. (1986). Salience of feedback on error and its effect on EFL writing quality. *TESOL Quarterly*, 20, 83-95.
- [23] Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics*, 11, 129-158.
- [24] Semke, H. (1984). The effects of the red pen. *Foreign Language Annals*, 17, 194-202.
- [25] Sheen, Y. (2007). The effect of focused written corrective feedback and language aptitude on ESL learners' acquisition of articles. *TESOL Quarterly*, 41, 255-283.
- [26] Shepard, K. (1992). Two feedback types: Do they make a difference? *RELC journal*, 23, 103-110.
- [27] Storch, N. (2007). Investigating the merits of pair work on a text editing task in ESL classes. *Language Teaching Research*, 11, 143-159.
- [28] Truscott, J. (1996). The case against grammar correction in writing classes. *Language Learning*, 46, 327-369.
- [29] Truscott, J. (1999). The case for "The case against grammar correction in L2 writing classes". A response to Ferris. *Journal of Second Language Writing*, 8, 111-122.
- [30] Truscott, J. (2007). The effect of error correction on learners' ability to write accurately. *Journal of second language Writing*, 16, 255-272.
- [31] Truscott, J., & Hsu, A.Y. p. (2008). Error correction, revision, and learning. *Journal of Second Language Writing*, 17, 292-305.
- [32] Van Beuningen, C., De long, N. H., & Kuiken, F. (2008). The effect of direct and incorrect corrective feedback on L2 learners' written accuracy. *ITL international Journal of Applied Linguistics*, 156, 279-296.
- [33] Van Beuningen, C., De long, N. H., & Kuiken, F. (2012). Evidence on the effectiveness of comprehensive error correction in Dutch multilingual classroom. *Language Learning*, 62, 1-41.

Fatemeh Esfandiari holds an M.A. in TEFL from Islamic Azad University, Ayatollah Amoli, Science and Research Branch (Amol, Iran, 2013). She has been an English instructor since (2003). Her main area of interest include Written Corrective Feedback, Pair Work interaction, Collaborative Dialogue and Teaching English to young adults.

Baqere Yaqubi is an assistant professor at Mazandaran University (Babolsar, Iran). Dr. Yaqubi received his PH.D. at University of Leeds (UK). His main areas of interest are interlanguage Pragmatic, Classroom discourse, Task-based Language Teaching and Learning, and Theories of Second Language Teaching and Learning.

Amir Marzban is an assistant professor at Islamic Azad University, Qaemshahr Branch, Iran. Dr Marzban obtained his Ph.D. from Islamic Azad University, Science and Research Branch, Tehran, Iran. His research interests are Applied Linguistic, ESP, CALL, and Materials Development.