Konin Language Studies
Faculty of Humanities and Social Sciences, State University of Applied Sciences in Konin, Poland
KSJ 9 (4). 2021. 435-454
http://ksj.pwsz.konin.edu.pl
doi: 10.30438/ksj.2021.9.4.2

The effects of written corrective feedback and revision on second language development

Saeed Roshan
Auckland University of Technology, New Zealand
https://orcid.org/0000-0003-3163-9803
sroshan@aut.ac.nz

Abstract

This article presents the findings of a study investigating the effectiveness of four types of written corrective feedback (i.e., direct corrective feedback, direct corrective feedback plus revision, metalinguistic explanation, metalinguistic explanation plus revision) on a complex linguistic structure, the English passive voice. The study was undertaken with 100 university students with a pre-test, treatment, immediate and delayed post-tests design. The findings showed that a single episode of written CF improved accuracy immediately and over time for all the experimental groups, but not for the control group. They also revealed that direct CF led to more improved accuracy compared to the other experimental groups. The results of this study contribute to our understanding of how written corrective feedback works and validate aspects of Gass's (1997) cognitive processing model and skill acquisition theories (DeKeyser, 2015) for the written context.

Keywords: written corrective feedback; revision; second language development

1. Introduction

The contribution of written CF to second language (L2) development has been the focus of a growing number of studies over the last 20 years. So far, written CF research has given much time to investigating whether a learner's accuracy improves after receiving written CF (e.g., Bitchener, 2008; Karimi & Fotovatnia, 2012;

Sheen, 2007; Van Beuningen, De Jong & Kuilken, 2008, 2012). However, the extent to which written CF plays a role in learners' acquisition of the target language (TL) is a problem that second language acquisition researchers and teachers of second language writing continue to face, an issue that has not been resolved and continues to be addressed. Thus, this paper aims to contribute to the written CF research in three respects. The first is to investigate the effect of written CF types, direct corrective feedback (DCF) and metalinguistic explanation (ME) on L2 development. In this study, written CF types have different degrees of explicitness, that is, DCF is more explicit than ME. In other words, direct CF learners who participated in this study had their errors identified and were explicitly provided with the correct forms, while the ME group was only provided with a handout that explicitly explained and exemplified the target structure (i.e., the English passive voice), meaning the errors were not identified for this group. The second objective is to explore the extent to which revision following feedback enhances the effect of written CF. There are few studies that have investigated this issue (e.g., Shintani, Ellis & Suzuki, 2014; Truscott & Hsu, 2008; Van Beuningen et al., 2012), and in each case the participants revised their text while they had access to the text on which they had received feedback. However, in this study, for the first time, the revision group revised their text without having access to the text on which they had received feedback. This is because revision following feedback, especially when learners have no access to the text on which they receive feedback, may lead to pushed output (Shintani et al., 2014). Swain (1985, 1995) argued that when learners are *pushed* to produce a language, they likely "notice a gap between what they want to say and what they can say, leading them to recognize what they do not know, or know only partially" (pp. 125-126). Swain takes the position that pushed output helps learners to notice grammatical forms that probably otherwise would go unattended. In addition, when learners revise their text and have access to the corrections (e.g., direct CF), little or no cognitive processing may be needed (Bitchener, 2016). Finally, this paper investigates the effectiveness of written CF in the acquisition of a complex structure, that is, the English passive voice. To date, most of the research has focused on simple linguistic structures such as English articles and the past tense. To the best of the author's knowledge, this study is the first to target the English passive voice as a complex structure.

2. Literature review

2.1. The relative effectiveness of different types of written CF

Written CF can be either unfocused or focused. *Unfocused* CF, or comprehensive feedback, is what many language teachers normally use in writing classes; that is, they generally provide feedback on all writing errors or at least on a range of errors. In

contrast, *focused* CF can be either "highly focused," that is, feedback is provided on only one error, or "less focused," that is, feedback is provided on a limited number of errors (Ellis et al., 2008). This study has chosen to explore highly focused feedback.

Ellis (2005), Gass (1997) and Schmidt (1994) emphasize the importance of *attention* and *understanding* in cognitive information processing. Thus, it would seem that if written CF is given on only one or a few linguistic errors (i.e., focused written CF), learners may be more likely to *attend* to the focused written CF feedback and *understand* the reason for the error and how to correct it. Even though unfocused written CF matches real teaching practice and therefore has high ecological validity (Van Beuningen et al., 2012), focused feedback may provide learners with a better opportunity to restructure their interlanguage system, as they receive repeated evidence to correct the same error (Shintani et al., 2014).

In this study, direct CF and ME as two types of written CF were compared in the treatment of the English passive voice. Direct CF was considered to be more explicit because it provided learners with the correct form of the erroneous form of the passive voice. However, ME, as used in this study, was considered to be less explicit because it was provided in the form of a handout to students that included an explanation and examples of the target structure, the passive voice, without identifying the errors in their writing text.

To the best knowledge of the present author, only two recent studies (i.e., Shintani et al., 2014; Shintani & Ellis 2013) have investigated the efficacy of direct written CF and ME in the form of a handout to students. Shintani and Ellis (2013) investigated the comparative effect of metalinguistic explanation and direct CF on 49 low-intermediate ESL learners' accurate use of the English indefinite article. They reported that even though the metalinguistic group outperformed the direct correction group in the immediate post-test, there was no difference in the efficacy of ME and direct correction for improving accuracy over time (after two weeks). However, Shintani et al. (2014) reported different findings. They compared the effects of direct CF and ME on 214 Japanese learners' accurate use of two grammatical structures: the hypothetical condition and the indefinite article. The researchers found that direct CF was more effective than ME in the delayed post-test (after two weeks). The authors suggested that the reason the findings differed from those of Shintani and Ellis (2013) may have been that, in the earlier study, there was a single structure, while in the second study the focus was on two structures, which may have led to an overload of information for learners at a lower proficiency level.

2.2. The effect of revision on accuracy

Theoretically, providing learners with an opportunity to revise a text plays an important role in the development process. Asking learners to perform revision

tasks may foster learning because "producing the correct form may help learners automatize their production" (Loewen, 2004, p. 157). When learners are revising their text with no access to the text on which they previously received feedback (as was the case in this study), they retrieve information from their long-term memory and this helps facilitate the consolidation of L2 knowledge. In addition, revision can also be argued from a skill-learning perspective because modifying output through revising and correcting an initial draft can provide the practice required for the proceduralization of explicit knowledge (Frear, 2012).

A number of studies (e.g., Truscott & Hsu, 2008; Van Beuningen et al., 2008, 2012) have examined whether revision results in improved accuracy in new texts. Truscott and Hsu (2008) reported that the increased accuracy revealed by their experimental group in the revision of their texts was not shown in their writing of new texts. However, the findings of Van Beuningen et al. in two different studies (2008, 2012) contradict the findings of Truscott and Hsu (2008). The main difference between Van Beuningen et al.'s (2008, 2012), and Truscott and Hsu's (2008) studies is the degree of explicitness of the written CF, and this may have led to the contradictory findings. In contrast to Truscott and Hsu's (2008) study in which one type of written CF with a low level of explicitness (underlining) was employed, in Van Beuningen et al.'s (2008, 2012) studies the researchers used two more explicit written CF types (direct correction and error codes). This may indicate that the degree of explicitness of written CF can impact the effectiveness of written CF. Therefore, these conflicting findings reveal a need for more studies to show if revising a text leads to improved accuracy in a new piece of writing. In addition, revision following feedback while learners have no access to the text on which they receive feedback has yet to be investigated. Thus, this is one of the issues the present study investigates.

2.3. The linguistic focus of written CF

Even though there is a growing body of evidence that written CF does improve accuracy over time, there has been limited research to investigate the effectiveness of written CF with specific linguistic errors involving complex linguistic structures. Furthermore, the theoretical arguments show that the extent to which written CF can effectively improve different types of linguistic errors may differ. The reason for this is that morphological, syntactic and lexical errors represent different domains of linguistic knowledge (Ellis, 2008; Ortega, 2009), and learners may need to focus their attention on more than one linguistic element each time they make a hypothesis on the use of the correct linguistic form for a certain linguistic error type.

A number of studies have indicated that written CF can improve the accuracy of simple rule-based categories such as English articles and the past tense

(e.g., Bitchener, 2008; Bitchener & Knoch, 2008; Frear, 2012; Sheen, 2007; Shintani & Ellis 2013). To the best of my knowledge, there are only two studies that have focused on a complex structure: Shintani et al. (2014) and Rummel (2014) targeted the hypothetical conditional and the present perfect tense respectively. Rummel (2014) found that written CF was effective for improving the accurate use of the present perfect tense immediately and over time (seven weeks). However, Shintani et al. (2014) reported that learners who received ME did not sustain improved accuracy in the use of the hypothetical conditional over two weeks. Thus, there is a need for more investigation before any generalised conclusion can be drawn. Therefore, to fill the gap, this study investigated the effectiveness of written CF in the use of the English passive voice as the target structure.

3.The study

The aim of this study was to examine the effect of different types of written corrective feedback (CF) (direct CF, metalinguistic explanation, direct CF plus revision, metalinguistic plus revision) on learners' output (immediate and delayed) in relation to the English passive voice. The research question addressed in this research was as follows:

RQ: What effect do focused direct corrective feedback (DCF) and metalinguistic explanation (ME) with and without revision have on learners' use of the English passive voice in an immediate text revision and in new texts over time?

3.1. Participants

The Departments of Foreign Languages and Literature at two universities in Iran were chosen as the research site for this study. A total of 135 L1 Persian male and female EFL learners at an intermediate level participated in this study. They were randomly divided into five groups encompassing four treatment groups and one control group. Each treatment group received one of the following CF types on their errors in using the English passive voice: direct written CF (DCF), direct written CF with revision (DCF+R), metalinguistic explanation (ME) and metalinguistic explanation with revision (ME+R). The control group (CN) only took the pre-test and post-tests and did not receive any CF treatment. After eliminating the participants who did not attend the post-tests phases of the study, the number of participants was 100: DCF (*N* = 20), DCF+R (*N* = 20), ME (*N* = 20), ME+R (*N* = 19), CN (*N* = 21).

3.2. Target structure

The target structure of this study was the English passive voice. In a passive construction, the object of an action takes the place of the subject of a sentence. The subject of the sentence (the agent or source of the action) may be omitted or added in a prepositional phrase. Passive and active constructions exist in both English and Persian. In the English language, the passive voice is formed by a combination of an auxiliary and the past participle of a transitive verb. For example *he closes the window* (active) becomes *The window is closed by him* (passive). In the Persian language, the passive voice is formed by a past participle and a derivation of the auxiliary verb „šodan," as in, for example, . سیبها چیده و شسته شده اند

Birjandi, Maftoon and Rahemi (2011) argued that the passive voice is particularly problematic for Iranian EFL learners because they have to apply a default processing strategy that ascribes the position of subject/agent to the first noun or phrase they hear or see in the input. Even though many Iranian EFL textbooks include a chapter about the passive voice, most EFL learners, even at advanced levels, incorrectly form the passive voice in their speaking and writing (Hinkel, 2002). Additionally, the participants of the main study were Iranian students who had academic writing and scientific reporting as two components of their study. The ability to use the passive voice could be helpful for the students when writing their assignments for the course. Most academic writing texts do not focus on who is the "doer" in an action, but on who/what is experiencing or receiving the action. Furthermore, it is irrelevant or repetitive in academic writing to state who is doing an action; therefore, the passive voice can be employed to construct these types of sentences. For these reasons, this study focused on the simple present form of the passive voice, as in *Diamonds are mined in South Africa.*

3.3. Writing tasks

Process tasks were utilized to elicit students' production of the passive voice. Such tasks deal with a number of stages that are in chronological order. Accordingly, it is necessary to start at the beginning and describe each stage through to the last one. Process tasks generally target the description of processes (how something happens) or procedures (how something is done) and often uses the passive structure (Swales & Feak, 2001). The three tasks given to students concerned how chocolate is produced, how apples are canned, and how coffee is produced. The participants were required to write approximately 200 words within 30 minutes for each task.

3.4. Design and treatment procedure

Four experimental groups (i.e., direct corrective feedback, direct corrective feedback and revision, metalinguistic explanation, metalinguistic explanation and revision) and one control group participated in five sessions (see Table 1). The treatments were operationalized as four different written CF strategies: direct written CF, direct CF and revision, written metalinguistic CF, and written metalinguistic explanation and revision. These are explained in more detail below.

Group 1 DCF

The learners produced writing task 1 (the pre-test) in Week 1. To do so, they received writing task 1 along with a sample answer, which they were required to read for 10 minutes. Then, the sample answer was collected and the participants started the writing task with 30 minutes to complete it. In Week 2, the direct CF group were provided with the written CF on the writing they had completed in the pre-test. The feedback was provided on the passive voice structure and the learners were given 10 minutes to look over the corrections to their writings. Then, these first texts with the feedback were collected and the group undertook task 2 (immediate post-test) over 30 minutes, using the same procedures as in the pre-test. In Week 4, the learners produced writing task 3 (the delayed post-test), following the same procedures as in the immediate post-test.

Group 2 DCF+ R

This group followed the same procedures as the DCF group in Week 1, but in Week 2, the learners were provided with 30 minutes to rewrite their first text. They were not allowed to look over their corrected text while rewriting. Then, the teachers collected the rewritten texts. Following that, the participants immediately wrote their second writing text (immediate post-test) following the same procedures as in the pre-tests. They wrote their fourth writing task in Week 4, following the same procedures as in the immediate post-test.

Group 3 ME

The metalinguistic explanation group wrote their first writing task using the same procedures as the last two groups in Week 1. In Week 2, the learners were not given any feedback on their writing in the pre-test. Instead, the teachers provided them with their first written text and a handout with an explicit explanation about the passive voice. The handout was in Farsi to prevent any lack of English vocabulary impacting their understanding. It included the definition of the English passive voice, instructions on how to make the English passive voice and some instances of

the English passive voice in both English and Farsi. The group was given 10 minutes to read the handout and check their text to see if/where there were passive voice errors. Then, the handouts and the written texts were collected by the teachers. Students immediately completed their second writing text (immediate post-test) using the same procedures as in the pre-test. In Week 4, the students completed their third writing task following the same procedure as in the immediate post-test.

Group 4 ME+R

The metalinguistic explanation plus revision group followed the same procedures as the ME group in Week 1. In Week 2, participants were provided with the writing task they had completed in Week 1 and the same handout that had been provided to the ME group (explicit explanation about the passive voice). The participants were given 10 minutes to read the handout. Then, the first text and the handout were collected and students were given 30 minutes to rewrite their first texts. Following that, the teachers collected the rewritten text. Participants then immediately wrote their second writing text (immediate post-test) following the same procedures as in the pre-test. The participants wrote their third writing task in Week 4, following the same procedures as in the immediate post-test.

Group 5 Control

The control group did not receive any feedback on their writing texts. Participants performed their writing tasks using the same procedures as the experimental groups in Week 1. They undertook the immediate post-test in Week 2 and the delayed post-test in Week 4. They completed each test in 30 minutes.

Table 1 Study design

	DC Group (N = 20)	DC+R Group (N = 20)	ME Group (N = 20)	ME+R Group (N = 19)	Control Group (N = 21)
Week 1	Written task, pre-test (Time 1)				
Week 2	DC (10 min)	DC (10 min) + Revision (30 min)	ME (10 min)	ME (10 min) + Revision (30 min)	No Treatment
	Written task, immediate post-test (Time 2)				
Week 4	Written task, delayed post-test (Time 4)				

Note. DCF = Direct corrective feedback, DCF+R = Direct corrective feedback and revision, ME = Metalinguistic explanation, ME+R = Metalinguistic explanation and revision

3.5. Scoring of writing tasks

The simple present passive voice is the target structure in this study. In the English language, the simple present passive voice is formed with a combination of "to be" and the past participle. Learners' use of the target structure can be scored in

two ways. In an absolute method, participants receive credit only if they use both of its components accurately. In other words, if they use only one out of two components accurately, they do not receive a score. The second method is a partial scoring system (Type 2), where each of the two components ("to be" and the past participle) is scored. It can be argued that a partial scoring method can provide a more accurate analysis of data than an absolute scoring method because the passive voice is made up of two components and students may acquire only one of its components instead of two as a result of feedback. Thus, they need to get credit for each correct component. The Type 2 method also provides a score for participants who attempt to apply the passive voice but use it inaccurately, because it shows they are in the process of learning the target structure. Literature shows that previous studies that have targeted structural problems (e.g., the hypothetical conditional) have used a Type 2 point system to analyse data (e.g., Shintani & Ellis, 2014).

To do so, if participants used the correct form of the passive voice, they received 2 points. That is, 1 point for the correct usage of "to be" and 1 point for the correct usage of the past participle. If only one of these was correct, they received only 1 point. If participants showed they were attempting to employ the passive voice, they were also awarded; that is, if participants applied the incorrect form of the verb "to be" (e.g., "are" was used instead of "is"), they were awarded .5 point, as they had attempted to use the verb "to be". If they used the wrong form of the past participle, (e.g., "spreaded" instead of "spread"), they were awarded .5 point (See Tables 3, 4 and 5). If they did not show any attempt, they did not receive any points, that is, zero (0) points. Correct sentences relating to Tables 2, 3 and 4 are the following: The door *is opened*. The letters *are written.* The table *is set*.

Table 2 Regular past participle

Error	Student answers	Mark
------	The door *is opened*	1+1
No *be*	The door *opened*	0+1
Wrong form of *be*	The door *are opened*	.5+1
No past participle	The door *is open*	1+0
No *be* /no past participle	The door *open*	0+0

Table 3 Irregular past participle

Type of error	Student answers	Mark
-------	The letters *are written*	1+1
No *be*	The letters *written*	0+1
Wrong form of *be*	The letters *is written*	.5+1
Wrong form of past participle	The letters *are writed*	1+.5
No past participle	The letters *are write*	1+0
No *be*/no past participle	The letters *write*	0+0

Table 4 Irregular past participle: No change in form

Type of error	Student answer	Mark
--------	The table *is set*.	1+1
Wrong form of past participle	The table *is seted.*	1+.5
Wrong form of *be*	The table *are set.*	.5+1

After scoring all the writing tasks, the following formula was utilized to calculate a total percentage score for each student:

$$\frac{\text{Number of points scored}}{\text{Number of points possible (i.e., number of passive uses x 2)}} \times 100$$

3.6. Data analysis

SPSS version 22 was employed to analyse the statistical data for the writing tasks. In order to address the research question, firstly descriptive statistics were performed. Then, a series of one-way ANOVAs were employed to test the comparative effects of the treatments between groups on the writing task. Repeated measure ANOVAs with pre- and post-test scores were conducted to measure within-group improvement of the experimental feedback groups and control group over time. In pairwise comparisons, both *p* value and effect size in the form of Cohen's *d* were used to evaluate the significance of the difference in mean scores. Cohen's *d* was used for pairwise comparisons and partial eta squared (η^2) was employed for ANOVAs. Effect sizes for the ANOVAs were calculated as eta-squared (η^2) with values of .138, .06 and .01 indicating large, moderate, and small effects respectively. Effect sizes for the pairwise-comparisons were estimated using the correlation coefficient r with values of .8, .5 or .2 indicating large, moderate, and small effects, respectively (Pallant, 2013).

The initial one-way ANOVAs detected significant effects in the pre-test for the research question. The significant effect of the pre-test could be a concern because if there were significant differences between groups in the pre-test, group differences in the immediate and delayed post-tests could be partially due to differences in the pre-test and not due to treatment. In order to control this concern, a one-way ANOVA with gain scores was conducted. In experimental studies, as covariate adjustment for a pre-test can cause biased results, it is recommended to employ gain scores because they are an unbiased estimate of true change (Rogosa, 1988). Thus, gain scores were performed to investigate learners' improvement from the pre-test to the post-tests for whole groups.

4. Findings

In order to address the research question posed for this study, firstly, descriptive statistics with raw scores for the accuracy scores of the English passive voice in

the written tasks were calculated. Table 6 shows the descriptive statistics for the four treatment groups at the three different testing periods – Time 1 (pre-test), Time 2 (immediate post-test) and Time 3 (delayed post-test). The mean scores refer to the mean percentage accuracy in partial obligatory occasions.

Table 5 Descriptive statistics for the accuracy scores in the written tasks

Group	*N*	Time 1		Time 2		Time3	
		M	*SD*	*M*	*SD*	*M*	*SD*
DCF	20	29.63	23.44	78.80	19.62	70.69	25.66
ME	20	45.83	28.01	79.86	16.93	80.21	18.70
DC+R	20	51.45	24.79	85.50	11.62	84.04	13.05
ME+R	19	55.20	26.65	85.55	12.07	84.50	11.31
CN	21	54.40	31.92	58.70	29.11	41.07	33.21

Note. DCF = direct corrective feedback with no revision, ME = metalinguistic explanation with no revision, DCF+R = direct corrective feedback with revision, ME+R = metalinguistic explanation with revision, CN = control group.

Table 5 shows that from the pre-test to the immediate post-test, the accuracy of all four experimental groups increased considerably; however, from the immediate post-test to the delayed post-test, the changes in the experimental groups were slight, that is, accuracy in the ME group slightly increased, while accuracy in the other groups slightly decreased. Accuracy in the control group also increased slightly from the pre-test to immediate post-test, but it then decreased sharply in the delayed post-test.

Overall, the findings showed that (a) experimental groups performed better in the post-tests than in the pre-test; (b) the experimental groups showed higher scores than the control group on both post-tests; (c) in the pre-test, the mean scores for direct CF (*M* = 29.63*, SD* = 23.44) were low in comparison to other groups – direct CF plus revision (*M* = 51.44*, SD* = 24.79), metalinguistic explanation (*M* = 45.83*, SD* = 28.01), metalinguistic explanation plus revision (*M* = 55.20*, SD* = 26.65) and control group (*M = 54.40, SD = 31.92*) (see Table 5). Due to the concerns that significant differences could be due to differences in Time 1 rather than the treatment, a one-way between groups ANOVA was employed to compare groups at Time 1.

On the pre-test (Time 1), the results of the one-way ANOVA showed significant differences among the pre-test scores of the five groups: $F(4, 95) = 3.067$, $p = .02$. The learners' pre-test scores were subjected to post hoc pairwise comparisons to investigate the comparative effects of feedback types on the passive voice structure. Post-hoc comparison using Bonferroni adjustments indicated that the direct CF group was significantly lower than metalinguistic plus revision and the control groups. Thus, the finding at Time 1 indicated that the possibility that group differences in Times 2 and 3 could be partially due to differences in Time 1 instead of the instructional treatments, which could not be precluded. As explained in the section

on statistical analyses, in order to control this concern, raw scores were changed to gain scores. Table 6 shows the descriptive statistics for the four treatment groups for the immediate gain (gain 1) and the delayed gain (gain 2). Gain 1 was calculated by subtracting immediate post-test scores from pre-test scores, and gain 2 was measured by subtracting delayed post-test scores from pre-test scores.

Table 6 Descriptive statistics of gain scores

Groups	N	Gain 1		Gain 2	
		M	*SD*	*M*	*SD*
DC	20	49.44	22.41	41.33	29.70
ME	20	34.03	28.63	34.38	23.94
DC+R	20	34.05	25.91	32.59	24.97
ME+R	19	30.35	23.19	29.30	22.25
Control	21	4.30	30.38	-13.32	32.77

Note. DCF = direct corrective feedback with no revision, ME = metalinguistic explanation with no revision, DCF+R = direct corrective feedback with revision, ME+R = metalinguistic explanation with revision, CN = control group.

In order to investigate the comparative effects of the treatments (between groups) on writing task scores, one-way ANOVAs and pairwise comparisons were conducted with gain scores. The results revealed that for gain 1, the result of a one-way ANOVA showed significant differences between groups: $F(4, 95) = 7.94$, $p < .001$, $\eta^2 = .25$.

Table 7 Effect sizes in the form of Cohen's d for the accuracy scores between groups

Group Contrast	Gains 1		Gains 2	
	d^1	p^2	*d*	*p*
DC vs. ME	0.59	.67	0.25	1.00
DC vs. DC+R	0.63	.68	0.31	1.00
DC vs. ME+R	0.83	.26	0.45	1.00
ME vs. ME+R	0.14	1.00	0.21	1.00
ME vs. DC+R	0.00	1.00	0.07	1.00
ME+R vs. DC+R	0.15	1.00	0.13	1.00
DC vs. CN	1.69	.00	1.74	.00
ME vs. CN	1.00	.005	1.66	.00
DC+R vs. CN	0.97	.005	1.57	.00
ME+R vs. CN	0.96	.02	1.52	.00

Note. 1. Effect size (Cohen's *d*). 2. Results of null hypothesis significance testing. ME = metalinguistic explanation with no revision, DC= direct corrective feedback with no revision, ME+R = metalinguistic explanation with revision, DC+R = direct corrective feedback with revision, CN = control group.

The findings reported in Table 7 show the resulting pairwise comparisons and Cohen's *d* values for the experimental and control groups. For gain 1, the findings indicated that the experimental groups significantly outperformed the CN group. Cohen's *d* values for direct CF, direct CF plus revision, metalinguistic explanation,

and metalinguistic explanation plus revision versus CN were 1.69, 0.97, 1.00, and 0.96 respectively, which in Cohen's terms would be considered large effect sizes.

Although all the experimental groups had large effect sizes, the value of direct CF was higher than in the case of the other groups. The analyses failed to detect significant differences between the experimental groups. Cohen's d values for all experimental groups versus each other ranged from small to large, that is, 0.00 to 0.83.

The findings of pairwise comparisons and Cohen's d values for gain 2 are shown in Table 7. The finding of a one-way ANOVA showed significant differences between groups: $F(4, 95) = 13.39$, $p < .001$, $\eta^2 = 0.36$. The results indicated that the experimental groups performed significantly better than the control group. Cohen's *d* values for direct CF, direct CF plus revision, metalinguistic explanation, and metalinguistic plus revision versus the control group were 1.74, 1.57, 1.66, and 1.52 respectively, which in Cohen's terms would be considered large effect sizes. The value of direct CF was the largest. There were no significant differences between experimental groups with small effect sizes. The values ranged from 0.07 to 0.45.

Within-groups ANOVA is used when the same participants are measured under different conditions (or measured at different points in time) (Pallant, 2013). Thus, within-groups ANOVA was chosen to compare scores on writing in the experimental feedback groups at Time 1 (pre-test), Time 2 (immediate post-test) and Time 3 (delayed post-test). To do so, repeated measure ANOVAs with pre- and post-test scores were conducted. The means and standard deviations for groups are presented in Table 6. The findings of repeated measure ANOVAs showed that experimental and control groups had a significant effect for time: $F(2, 18) = 65.90$, $p < .001$, $\eta^2 = .58$. Additionally, the results of repeated measure ANOVAs for treatment and control groups revealed significant differences for each group with direct CF group: $F(2, 18) = 46.25$, $p < .001$, $\eta^2 = .83$; direct CF plus revision group: $F(2, 18) = 17.33$, $p < .001$, $\eta^2 = .65$; metalinguistic explanation group: $F(2, 18) = 20.55$, $p < .001$, $\eta^2 = .69$; metalinguistic explanation plus revision group: $F(2, 17) = 15.73$, $p < .001$, $\eta^2 = .64$ and control group: $F(2, 19) = 5.62$, $p = .01$, $\eta^2 = .37$.

Table 8 Effect sizes in the form of Cohen's d for the accuracy scores over time

Group	*N*	Time 1–Time 2		Time 2–Time 3		Time 1–Time3	
		d[1]	*p*[2]	*d*	*p*	*d*	*p*
DCF	20	2.28	.00	.35	.35	1.68	.00
ME	20	1.47	.00	0.01	1.00	1.44	.00
DC+R	20	1.75	.00	0.11	1.00	1.64	.00
ME+R	19	1.46	.00	0.08	1.00	1.43	.00
CN	21	0.14	1.00	0.56	.008	0.40	.23

Note. 1. Effect size (Cohen's d). 2. Results of null hypothesis significance testing. ME = metalinguistic explanation with no revision, DC = direct corrective feedback with no revision, ME+R = metalinguistic explanation with revision, DC+R = direct corrective feedback with revision, CN = control group

The findings of pairwise comparisons (paired *t*-test) and the within-group effect sizes for experimental and control groups are reported in Table 8. The findings revealed that all the experimental groups significantly improved from Time 1 to Time 2. Cohen's *d* values for direct CF, direct CF plus revision, metalinguistic explanation, and metalinguistic explanation plus revision from Time 1 to Time 2 were 2.28, 1.75, 1.47, and 1.46 respectively, which in Cohen's terms would be considered large effect sizes. The value of direct CF was the largest. There were no significant differences for all groups from Time 2 to Time 3, as small effect sizes were found, with values ranging from 0.01 to 0.35. That indicates the experimental groups were able to sustain this improvement. However, the mean scores at Time 3 were still significantly greater than those of Time 1 with large effect sizes. Cohen's *d* values for direct CF, direct CF plus revision, metalinguistic explanation, and metalinguistic explanation plus revision from Time 1 to Time 3 were 1.68, 1.64, 1.44, and 1.43 respectively, which in Cohen's terms would be considered large effect sizes. The direct CF group had the biggest value. The control group showed no significant differences from Time 1 to Time 2 with small effect size (i.e., 0.14); however, the scores significantly decreased from Time 2 to Time 3 with medium effect size (i.e., 0.56). There was no significant difference from Time 1 to Time 3 with small effect size (i.e., 0.40). In summary, all experimental groups showed the effectiveness of corrective feedback over time. However, the efficacy of the direct CF group proved to be relatively higher in both the short and long terms than other groups.

5. Discussion

The results of a series of ANOVAs showed that the experimental groups (direct CF, direct CF plus revision, metalinguistic explanation, metalinguistic explanation plus revision) significantly improved their accuracy from the pre-test to the immediate post-test. Subsequently, from the immediate to the delayed post-test the improvement deteriorated slightly, but the decrease in accuracy was not statistically significant. This reveals that some learners retained the improvement from the immediate to the delayed post-test. Furthermore, the accuracy in the delayed post-test was significantly higher than that of the pre-test. Additionally, the accuracy rate for the control group revealed no significant improvement from the pre-test to the immediate post-test and from the immediate post-test to the delayed post-test. These findings not only support the theoretical expectations but also confirm the results of previous studies.

The results of the between group analysis showed that all experimental groups (direct CF, direct CF plus revision, metalinguistic explanation, metalinguistic explanation plus revision) outperformed the control group in both the

short term and over four weeks. There is a theoretical explanation for why written CF can be considered to play a facilitative role in L2 development. Gass's (1997) framework identifies how a single episode of input processing (e.g., in the form of written CF) may help learners to develop their explicit knowledge. Accordingly, if noticed and comprehended, input can subsequently go through central processing (i.e., intake and integration), and result in output. Thus, the findings of the current study have shown that a single written CF treatment can help learners to improve the accuracy of a complex structure such as the English passive voice, and that accuracy is not only evident in the short term but also in the writing of a new text after a period of time.

The findings showed that direct CF treatment was more effective than the other treatments in both the short term and over time, and also from the pre-test to the immediate and delayed post-tests (Bitchener, 2008; Shintani et al., 2014). This can be explained theoretically in that the degree of explicitness provided to the direct CF group was higher than that of the ME group. In other words, in this study, direct CF was more explicit because the direct CF group was provided with the correct form of the target structure, the English passive voice. On the other hand, ME was less explicit because the feedback for the ME group was in the form of a one-page handout without errors in the use of the English passive voice being identified in the students' written text. Thus, the ME group received no explicit feedback on their pre-test writings. Learners' existing knowledge and the salience of the input play important roles in the processing of the input, and more explicit types of written CF are likely to draw learners' attention to a greater extent than less explicit written CF types (e.g., metalinguistic explanation, as used in this study). More explicit types of feedback (e.g., direct CF) may also reduce the confusion that learners may experience if they do not understand less explicit types of CF. This is because more explicit types of feedback reveal learners' errors and may help learners to more clearly and fully comprehend their errors than less explicit CF. Additionally, more explicit types of written CF contain more linguistic information, which may lead to the formation of a new hypothesis about the target structure (e.g., the English passive voice) and the production of output. Thus, because the degree of explicitness of written CF types may impact the level of attention learners pay to the input, and because the amount of linguistic information included in the input facilitates the formation of a new hypothesis, a more explicit type of written CF (e.g., direct CF) is more likely to lead to corrected output.

Empirically, the findings of the current study regarding the high level of effectiveness of direct CF in comparison to other types of written CF are supported by the results of earlier studies. Bitchener (2008) reported that providing only direct CF was more effective than providing direct CF plus metalinguistic

explanation over two months. He argued that the possible reason was that the limited details of written ME may have been insufficient to result in a significant effect. Shintani et al. (2014) also reported that direct CF was more effective than metalinguistic explanation provided in the form of a handout in the immediate post-test. The reason was that the direct CF was more explicit and provided participants with the correct form of the target structure and thus enabled them to make a comparison between the input and their own erroneous structure.

In contrast to the present study, in her study of Laotian EFL learners, Rummel (2014) found no difference between the direct error correction group and the metalinguistic explanation group in treating target structures (simple past tense and present perfect tense). A possible reason for the difference in findings is that the metalinguistic explanation in the present study was in the form of a handout and the errors in the target structure were not identified on the students' writing. However, in Rummel's (2014) study the metalinguistic group received explicit feedback in the form of identifying the errors and providing explicit explanations on them. Thus, in the present study, direct CF and metalinguistic explanation feedback had a different degree of explicitness, that is, the direct CF was more explicit than the metalinguistic explanation in the form of a one-page handout. The direct CF and metalinguistic CF in Rummel's (2014) study were both explicit CF.

6. Conclusion

This study investigated the efficacy of direct CF and metalinguistic feedback, with and without revision, on learners' use of the English passive voice in a text written immediately after the provision of written CF and in new texts over time (i.e., four weeks). A unique feature of the present study was that for the first time, the English passive voice was targeted and the revision group had no access to the text on which they had received feedback. The findings showed that a single episode of written CF improved accuracy immediately and over time (after two weeks) for all the experimental groups, but not for the control group. They also revealed that direct CF led to more improved accuracy compared to the other groups.

The findings of this study validate aspects of Gass's (1997) cognitive processing model for the written context. In other words, Gass's (1997) framework identifies how a single episode of explicit input processing may help learners to develop their explicit knowledge. Accordingly, if noticed and comprehended, explicit input can subsequently go through central processing (i.e., intake and integration) and result in output. These findings validate skill acquisition theories for the written context. These theories posit that declarative knowledge must be processed with conscious attention (i.e., controlled processing) (McLaughlin, 1987) and that declarative knowledge must be proceduralized (Anderson, 2000).

Skill acquisition theories (DeKeyser, 2015) maintain that intentional learning (e.g., by means of explicit written CF) can play an important role in the controlled stage, when learners receive explicit input and draw on their procedural knowledge, and that such learning can ultimately result in more automatic processing of their procedural knowledge and, ultimately, their automatic knowledge. Thus, skill acquisition theories can be used to explain how learners may improve their L2 development of a target structure (e.g., the English passive voice) immediately and over time. Additionally, the absence of improvement by the control group in this study suggests that the improved accuracy of the treatment groups was the result of the explicit written CF provided.

Even though the present study has achieved its aims, a number of limitations and shortcomings are acknowledged. The first limitation concerns the use of the same type of writing tasks, that is, reconstruction tasks, in the pre-test, immediate post-test and delayed post-test. The nature of the reconstruction tasks may have created a context for the participants to use the target structure more accurately than in free writing, as the requirement to reconstruct a text could act as scaffolding for the learners. Second, the duration of this research may also be seen as a limitation. The study was conducted over four weeks and the delayed post-test was conducted two weeks after the immediate post-test. The findings showed statistically significant improved accuracy for the experimental groups over four weeks. However, it is unclear whether learners would have been able to retain this accuracy over a longer period of time. Thus, it is recommended that delayed post-tests be conducted after a longer time period than the two weeks that was used in the present study, for example, after 6 months.

References

Anderson, J. R. (2000). *Cognitive psychology and its implications* (5th ed.). Worth Publishers.

Birjandi, P., Maftoon, P., & Rahemi, J. (2011). VanPatten's processing instruction: Links to the acquisition of the English passive structure by Iranian EFL learners. *European Journal of Scientific Research, 64*(4), 598-609.

Bitchener, J. (2008). Evidence in support of written corrective feedback. *Journal of Second Language Writing, 17*(2), 102-118. doi: 10.1016/j.jslw.2007.11.004

Bitchener, J. (2016). To what extent has the published written CF research aided our understanding of its potential for L2 development? *ITL International Journal of Applied Linguistics, 167*(2), 111-131. doi: 10.1075/itl.167.2.01 bit.

Bitchener, J., & Knoch, U. (2008). The value of written corrective feedback for migrant and international students. *Language Teaching Research*, *12*(3), 409-431.

Bitchener, J., & Knoch, U. (2010). The contribution of written corrective feedback to language development: A ten-month investigation. *Applied Linguistics, 31*(2), 193-214. doi: 10.1093/applin/amp016

Bitchener, J., & Storch, N. (2016). *Written corrective feedback for L2 development*. Multilingual Matters.

Bruton, A., (2009). Designing research into the effect of error correction in L2 writing: not so straightforward. *Journal of Second Language Writing*, *18*(2), 136-140. https://doi.org/10.1016/j.system.2010.07.001

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 (3), 267-296. http://dx.doi.org/10.1016/S1060-3743(03)00038-9

DeKeyser, R. (2015). Skill acquisition theory. In B. VanPatten, & J. Williams (Eds.). *Theories in second language acquisition: An introduction* (pp. 94-122). Routledge.

DeKeyser, R. (Ed.). (2007). *Practice in a second language: Perspectives from applied linguistics and cognitive psychology*. New York: Cambridge University Press.

DeKeyser, R. (1995). Learning second language grammar rules: An experiment with a miniature linguistic system. *Studies in Second Language Acquisition* 17(03), 379-410.

Ellis, N. C. (2005). At the interface: Dynamic interactions of explicit and implicit language knowledge. *Studies in second language acquisition, 27*(2), 305-352. https://doi.org/ 10.1017 /S027226310505014X

Ellis, R. (2008). *The study of second language acquisition* (2nd ed.). Oxford University Press.

Ellis, R. (2009). A typology of written corrective feedback types. *ELT Journal, 63*(2), 97-107. doi:10.1093/elt/ccn023

Ellis, R., Sheen, Y., Murakami, M., & Takashima, H. (2008). The effects of focused and unfocused written corrective feedback in an English as a foreign

language context. *System*, *36*(3), 353-371. https://doi.org/10.1016 /j.system.2008.02.001

Ferris, D. R. (1999). The case for grammar correction in L2 writing classes: A response to Truscott. *Journal of Second Language Writing*, *8*(1), 1-11. https://doi.org/10.1016/S1060-3743(99)80110-6

Ferris, D. R. (2006). Does error feedback help student writers? New evidence on the short- and long-term effects of written error correction. In K. Hyland (Ed.), *Feedback in second language writing: Contexts and issues* (pp. 81-104). Cambridge University Press.

Frear, D. (2012). *The effect of written corrective feedback and revision on intermediate Chines learners' acquisition of English*. (Unpublished doctoral dissertation). University of Auckland, Auckland, New Zealand.

Gass, S. M. (1997). *Input, interaction and the second language learner*. Lawrence Erlbaum Associates.

Hinkel, E. (2002). *Second language writers' text: Linguistic and rhetorical features*. Routledge.

Karimi, M. & Fotovatnia, Z. (2012). The effects of focused vs. unfocused written teacher correction on the grammatical accuracy of Iranian EFL undergraduates. *Asian EFL Journal: Professional Teaching Articles – CEBU Issue, 62,* 42-59.

Loewen, S. (2004). Uptake in incidental focus on form in meaning-focused ESL lessons. *Language Learning*, *54*(1), 153-188. https://doi.org/10.1111/j.1467-9922.2004.00251.x.

McLaughlin, B. (1987) *Theories of second-language learning.* Edward Arnold.

Ortega, L. (2009). *Understanding second language acquisition*. Hodder Arnold.

Pallant, J. (2013). *SPSS survival manual: A step by step guide to data analysis using IBM SPSS* (4th ed.). Allen & Unwin.

Polio, C. (2012). The relevance of second language acquisition theory to the written error correction debate. *Journal of Second Language Writing*, *21*, 375-389.

Rogosa, D. (1988). Myths about longitudinal research. In K. W. Schaie, R. T. Campbell, W. M. Meredith, & S. C. Rawlings (Eds.), *Methodological issues in aging research* (pp. 171-209). Springer.

Rummel, S. (2014). S*tudent and teacher beliefs about written CF and the effect those beliefs have on uptake: A multiple case study of Laos and Kuwait*. (Unpublished doctoral dissertation). AUT University, Auckland, New Zealand.

Schmidt, R. (1994). Implicit learning and cognitive unconscious: Of artificial grammars and SLA. In. N. Ellis (Ed.), *Implicit and explicit learning of language* (pp.165-209). Academic Press.

Sheen, Y. (2007). The effect of focused written corrective feedback and language aptitude on ESL learners' acquisition of articles. *TESOL Quarterly*, *41*(2), 255-283. doi: 10.1002/j.1545-7249.2007.tb00059.

Shintani, N. (2017). The effects of the timing of isolated FFI on the explicit knowledge and written accuracy of learners with different prior knowledge of the linguistic target. *Studies in Second Language Acquisition*, *39*(1), 129-166. doi:10.1017/S0272263116000127

Shintani, N., & Ellis, R. (2013). The comparative effect of metalinguistic explanation and direct written corrective feedback on learners' explicit and implicit knowledge of the English indefinite article. *Journal of Second Language Writing*, *23*(3), 286-306.doi: 10. 1016/ j. jslw. 2013.03.011

Shintani, N., Ellis, R., & Suzuki, W. (2014). Effects of written feedback and revision on learners' accuracy in using two English grammatical structures. *Language Learning*, *64*(1), 103-131. doi:10.1111/lang.12029

Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. Gass, & C. Madden (Eds.), *Input in second language acquisition* (pp. 235-253). Newbury House.

Swain, M. (1995). Three functions of output in second language learning. In G. Cook & B. Seidlhofer (Eds), *Principle and practice in applied linguistics: Studies in honour of H. G. Widdowson* (pp. 125-144). Oxford University Press.

Swales, J. M., & Feak, C. B. (1994). *Academic writing for graduate students. A course for NNS of English.* The University of Michigan Press

Truscott, J. (1996). The case against grammar correction in L2 writing classes. *Language learning*, *46*(2), 327-369. doi:10.1111/j.1467-1770.1996. tb0 12 38.x

Truscott, J. (2010). Some thoughts on Anthony Bruton's critique of the correction debate. *System*, *38*(2), 329-335. https://doi.org/10.1016/j.system.2010.03.014.

Truscott, J., & Hsu, A. Y. p. (2008). Error correction, revision, and learning. *Journal of Second Language Writing*, *17*(4), 292-305. Doi: 10.1016/j.jslw.2007.05.003.

Van Beuningen, C. G., De Jong, N. H., & Kuiken, F. (2008). The effect of direct and indirect corrective feedback on L2 learners' written accuracy. *International Review of Applied Linguistics*, *156*, 279-296. doi: 10.1075/itl.156.24beu

Van Beuningen, C. G., De Jong, N. H., & Kuiken, F. (2012). Evidence on the effectiveness of comprehensive error correction in second language writing. *Language Learning*, *62*(1), 1-41. https://doi.org/10.1111/j.1467-9922.2011.00674.

Yang, Y., & Lyster, R. (2010). Effects of form-practice and feedback on Chinese EFL learners' acquisition of regular and irregular past tense forms. *Studies in Second Language Acquisition*, *32*(2), 235-263. https://doi.org/10.1017/S0272263109990519