The Effect of Sustained Teacher Feedback on CAF, Content and Organization in EFL Writing

Ali Rastgou [a, *], Neomy Storch [b], Ute Knoch [b]

[a] *Islamic Azad University, Iran*

[b] *University of Melbourne, Australia*

A B S T R A C T

Despite teachers' mainstream practices in L2 writing classrooms addressing different dimensions of writing over time, much of the research on feedback in recent years has been of relative short duration and has mainly focused on accuracy. The current longitudinal study investigated the influence of sustained teacher written feedback on accuracy, syntactic complexity, fluency, content, and organization in an EFL context. Ninety-two learners were divided into four groups, receiving written corrective feedback, feedback on content and organization, multilateral feedback (i.e., on grammatical accuracy, content, and organization), and no feedback over a 3½-month period. They completed a pre-test, post-test, and delayed post-test and wrote and revised eight interim expository compositions on a weekly basis. Results showed that the three treatment groups significantly improved in the dimensions on which they received feedback. However, only the groups who received feedback on content and organization improved in fluency. Importantly, the multilateral group improved in accuracy as well as fluency, content and organization. Theoretically, the findings endorse the language learning potentials of sustained writing as long as it is guided by teacher feedback. The findings provide empirical support for the influence of sustained feedback on expanding and consolidating learners' explicit knowledge of L2 writing.

Keywords: sustained feedback; CF; feedback on content and organization; multilateral feedback; accuracy; fluency; L2 writing development

© Urmia University Press

A R T I C L E   H I S T O R Y

* Corresponding author: Shahrood Branch, Islamic Azad University, Shahrood, Iran
Email address: ali.r@alumni.unimelb.edu.au

Introduction

The impact that teacher feedback has on the writing of second language (L2) learners is a topic that has attracted much research interest over the past two decades. However, one of the major disparities between the focus of current research on written feedback and teachers' feedback practices in L2 classrooms is aspects of writing on which feedback is provided. Whereas L2 writing practitioners tend to offer feedback on various aspects of writing including accuracy, content and organization, quantitative research has focused predominantly on feedback on grammatical accuracy. Studies that have considered the impact of feedback on content and accuracy (e.g. Hartshorn, Evans, Merrill, Sudweeks, Strong-Krause & Anderson, 2010; Manchón & Roca de Larios, 2011) have not included a control group. As a number of scholars have argued, the influence of feedback needs to be assessed not only on new compositions but also by comparison to a group that receives no feedback (Ferris, 1999; Truscott, 1996). Many of the more recent studies have investigated the influence of corrective feedback (CF), partly due to the importance attributed to grammatical accuracy in L2 writing, and partly to the debate in the literature on the efficacy of CF, sparked by Truscott (1996) and subsequently countered by many others (e.g., Bitchener, 2008; 2012; Ferris, 2010; Hyland, 2003; 2011; Manchón, 2011a).

This paper investigates the influence of feedback on different aspects of writing, using a range of quantitative measures. It also investigates whether feedback on grammatical accuracy, feedback on content and organization, and a combination of feedback on accuracy, content and organization lead to improved complexity, accuracy and fluency (CAF) in writing. We pinpoint that L2 written feedback studies need to encompass sustained engagement in writing, where students produce a number of written texts (DeKeyser, 1997) and receive frequent feedback treatments because L2 writing tends to take time to develop. Such a research design bears a closer resemblance to what L2 writing practitioners often do in L2 writing classrooms, which is likely to improve the external validity of the findings.

Literature review

The growing interest in written CF has produced over the past 20 years or so a very large body of research as well as some meta-analyses (e.g. Kao & Wible, 2014; Kang & Han, 2015; Russell & Spada, 2006). Our literature review deals with two strands that are of most relevance to our study. First, it begins with a review of studies investigating the impact of feedback on writing, where feedback was provided to L2 learners on content and/or grammatical accuracy. It then reviews research where feedback was provided on language use only and the impact of CF on grammatical accuracy.

Feedback on content along with/without CF

A number of studies have investigated the influence of content-related feedback in L2 writing. Some early longitudinal studies compared the impact of feedback on content and accuracy. Typically in these studies one group received either text-specific or generic feedback on content (Kepner, 1991; Sheppard, 1992) or combined feedback on content and CF (Semke, 1984). In these studies, feedback on content refers to comments and questions in response to content (Semke, 1984), general requests for clarification (Sheppard, 1992), and a summary of the students' main points, followed by comments addressing extending the topic (Kepner, 1991). None reported that the CF or feedback on other dimensions had an impact on accuracy, whereas all reported positive effects of content-related feedback on ideational quality and critical thinking (Kepner, 1991), syntactic complexity (Sheppard, 1992), and fluency (Semke, 1984). In Manchón and Roca de Larios' (2011) study, in which learners received feedback on all aspects of writing,

similar results were reported. Holistic analyses showed that the learners improved in communication, organization and argumentation, but not significantly in accuracy. However, none of these studies had a control group to ensure that their findings were guided by feedback and not by engagement in writing (Truscott, 1996). Furthermore, in a number of these studies (e.g. Kepner, 1991; Semke, 1984) students were not asked to revise their drafts. Students are less likely to pay attention to teacher feedback unless they are required to revise their compositions in response to feedback (Shintani, Ellis & Suzuki, 2014).

A study that attempted to address some of these shortcomings is by Hartshorn et al. (2010). This longitudinal study compared the impact of dynamic written CF and feedback on content and organization. While one group wrote and rewrote daily 10-minute compositions and received only CF, the other group was taught process writing and wrote four multi-draft papers while receiving CF and feedback on relevance, organization and coherence, and development of idea. The study reported that the CF-only group significantly improved in accuracy, without reduction in syntactic complexity and fluency. However, this study did not report the outcome of content- and organization-related feedback on students' writing.

A cluster of studies, known as revision studies, (e.g., Ashwell, 2000; Ferris, 1997) investigated the impact of offering feedback on form and content across two or three drafts of a single composition. Ashwell (2000) found that his treatment groups, who received three feedback types – content then form; form then content; and mixed – outperformed the no-feedback group in accuracy, whereas Ferris (1997) found that the students – all in one group – successfully incorporated the direct text-specific, content-related feedback in their subsequent drafts. Although the feedback approach in the revision studies resembled those in most authentic L2 writing classrooms, i.e., feedback was offered on both form and content, the feedback effect was only gauged in the revised drafts. Truscott (1996), who was particularly critical of the value of CF, argued that improvement in revised versions of a single composition does not signify L2 writing development or acquisition of new syntactic or morphological forms, and, that instead, effects of CF should be assessed on new pieces of writing. A number of researchers thus set out to investigate the impact of CF on new pieces of writing.

Feedback on grammatical accuracy

Studies on the impact of CF on L2 writing have tended to investigate whether targeting a restricted number of errors or providing direct versus indirect feedback may be best for improving L2 writing accuracy. One group of studies – focused CF studies – have investigated the impact of CF only on one or two grammatical structures. In these studies, the learners received direct feedback (correct forms provided) – with or without oral/written metalinguistic explanations – on the use of the targeted structures. Some findings show that, after receiving one-off treatments, the treatment groups outperformed control groups in the correct use of the targeted structures such as English indefinite article (e.g., Bitchener, 2008; Bitchener & Knoch, 2010; Sheen, 2007), simple past (Rummel & Bitchener, 2015), past perfect (Suzuki, Nassaji & Sato, 2019) and the hypothetical conditional (Shintani et al., 2014). Nevertheless, other studies show that the treatment groups did not outperform the control group in the correct use of articles (Pashazadeh, 2017; Shintani & Ellis, 2013; Shintani, et al., 2014; Suzuki, et al., 2019) and the infinitive and the hypothetical conditional (Pashazadeh, 2017). Although some scholars (e.g., Lee, 2019) argue for the greater benefits of adopting a focused approach for both students and teachers, Xu (2009) maintains that exclusive attention to a minimal number of grammatical structures can result in conscious attention of experimental groups to the target structure(s) in the post-tests, hence the lower generalizability of these findings (See also Bruton, 2009a; Storch, 2010).

Some studies (e.g., Ellis, Sheen, Murakami, & Takashima, 2008; Frear & Chiu, 2015; Sheen, Wright, & Moldawa, 2009) compared the effects of focused and unfocused CF on accuracy. The findings of these studies tend to support the efficacy of focused over unfocused feedback, with treatment groups outperforming control groups. However, there were some variations in these studies in terms of how many types of errors constitute unfocused feedback (e.g., see Sheen et al., 2009). Another limitation of these studies was the limited number of compositions the participants wrote without redrafting, which does not allow for deep engagement with feedback provided.

Other written CF studies explored the effects of unfocused written CF in one-off treatments (e.g., Truscott & Hsu, 2008; Van Beuningen, De Jong, & Kuiken, 2012). Van Beuningen et al. (2012) carefully designed study investigated the effects of direct and indirect CF on written accuracy, complexity and lexical diversity. The findings demonstrated the efficacy of both direct and indirect CF for accuracy improvement, without detriment to complexity and lexical diversity. However, one of the shortcomings of this study was the short duration of the treatment. Expecting L2 writers to notice their errors in numerous structures and generalize their knowledge following limited treatments runs counter to SLA theories (e.g., DeKeyser, 1997; Gass, 2003). Some longitudinal studies compared the influence of unfocused direct and indirect CF on accuracy (e.g., Chandler, 2003) or the influence of unfocused CF on accuracy (e.g., Ferris, 2006; Riazantseva, 2012). These studies, too, produced mixed results partly due to their design differences. None included a no-feedback group, and few required (all) their groups to rewrite their compositions.

Soleimani, Modirkhamene and Sadeghi (2017), however, investigated differences in grammatical accuracy, complexity and fluency between compositions written by collaborative and individual writers of intermediate and advanced proficiency level and found that both intermediate and advanced collaborative groups outperformed the individual writers in accuracy (error-free T-units and clauses) and fluency (average number of words, T-units and clauses per composition) across drafts after students rewrote their compositions guided by teacher CF, but not in complexity, predictably due to the feedback focus on grammatical accuracy.

In sum, the findings in most written feedback studies suggest that content-related feedback, whether offered exclusively or in tandem with CF, can lead to improvement in aspects of meaning, fluency or complexity, but not overall accuracy. Studies that investigated the efficacy of CF only, were more carefully designed, but did not yield conclusive findings concerning the efficacy of CF, particularly when adopting an unfocused approach.

The current study thus set out to investigate the impact of feedback on grammatical accuracy, content and organization. Our study was longitudinal, offering learners opportunities to receive sustained feedback, notice the feedback and engage with it by redrafting. These key elements accord with cognitive theories/hypotheses of second language acquisition, specifically the noticing hypothesis (Schmidt, 1990) and the skill acquisition theory (DeKeyser, 1997). The role of the two theories in L2 development was best noted by the weak interface theory of second language instruction (Ellis, 2011) proposing that explicit processing of knowledge plays a role in SLA through noticing the gap and guided output practice.

Schmidt's (1990) noticing hypothesis holds that language needs to be consciously noticed for learning to take place. Feedback may bring the salience of linguistic gaps to L2 learners' attention (Bitchener, 2012; Storch, 2010). Further, Schmidt (1990) argues that L2 learning goes beyond noticing the gap to the "generalization from instances, hypothesis formation, or the induction of linguistic rules", which "requires insight and understanding" (p. 145). It is conceivable that responding to feedback on a number of occasions may therefore encourage L2 learners to extrapolate patterns and rules. Furthermore, according to DeKeyser's (1997) skill acquisition

theory, learners need repeated practice to consolidate newly learnt knowledge. Before the new knowledge is consolidated, the learner's performance is characterized by noticeable error rate, long reaction time, and variability in use. The reciprocal role of noticing and prolonged engagement in L2 development was underscored by SLA scholars (N. Ellis, 1994; 2011; R. Ellis, 1994; 2009). They argued that explicit processing of knowledge plays a part in SLA by facilitating noticing the gap and guided output practice, and that sustained conscious output practice guided by teacher's explicit feedback may then have a role in promoting implicit knowledge and proceduralization (N. Ellis, 2011). DeKeyser (2007), however, admits that much of the research on practice has focused on the development of oral skills. Thus, we sought to investigate the role of sustained practice in writing, guided by teacher feedback, in writing development. Our notion of writing development includes evidence of greater ability to compose texts that contain relevant and well-supported ideas (content), as well as well-structured paragraphs and a more cohesive and coherent text (organization), and language that is not only more accurate but also more fluent and complex. The current study explored whether sustained feedback on grammar, content and organization versus mere engagement in sustained writing practice led to improvement in grammatical accuracy, syntactic complexity, fluency, content, and organization in compositions written by intermediate students in an EFL context. This multifaceted aim was operationalized as five research questions.

RQ1. Does sustained teacher written feedback facilitate the development of accuracy in L2 writing?

RQ2. Does sustained teacher written feedback facilitate the development of syntactic complexity and fluency in L2 writing?

RQ3. Does sustained teacher written feedback facilitate the development of content in L2 writing?

RQ4. Does sustained teacher written feedback facilitate the development of organization in L2 writing?

RQ5. To what extent does providing sustained multilateral feedback facilitate balanced L2 writing development?

Method

Context and participants

The current study was undertaken in a language institute in Shahrood, Iran, over three consecutive terms. The course book taught at the time focused mainly on developing speaking and listening skills. In the first week of the institute calendar, the first researcher attended target classes – Levels 8-13 selected from Level 1-16 classes – to introduce the study and invite students to participate in the writing study. (Level 1 and 16 represent low elementary and high intermediate, respectively, with respect to students' proficiency in English.) The participants were required to attend a writing session on a weekly basis to write a composition.

Of all volunteers, 92 students (63 females and 29 males) participated in all 11 sessions. Over 50% of the participants were high school students. Others were university students and graduates, ranging in age from late teens to early 30s. All participants were Persian speakers who had received up to seven years of formal English instruction at school. Based on the institutes' placement test, these learners were deemed low-intermediate to mid-intermediate. The participants were then randomly allocated to one of four groups (see details below). In doing so,

we attempted to ensure that the groups were similar in terms of age, gender and proficiency distribution.

Tasks and tests

Nine expository writing prompts adapted from TOEFL PBT writing topics were used. The same topic was used for the pre-test (Session 1), post-test (Session 10), and delayed post-test (Session 11) compositions: "Do you agree or disagree with the following statement? Face-to-face communication is better than other types of communication, such as letters, email, or telephone calls. Use specific reasons and examples to support your opinion." Eight other prompts – e.g., the influence of watching TV on kids, living in the countryside, eating home-made and restaurant meals, watching advertisements, and so forth – were used for interim sessions 2 to 9.

Group assignment

Following Week 1, each class was randomly assigned to one of the four groups. During Weeks 2 to 9, the three treatment groups received sustained feedback from the first researcher on their compositions, while the control group engaged in writing on the same topics without receiving feedback.

Group 1 received direct written CF (WCF). Direct CF incorporated insertions of correct forms where they were omitted or above erroneous forms that were crossed out. Excerpt 1 shows a sample of the CF given to the WCF group.
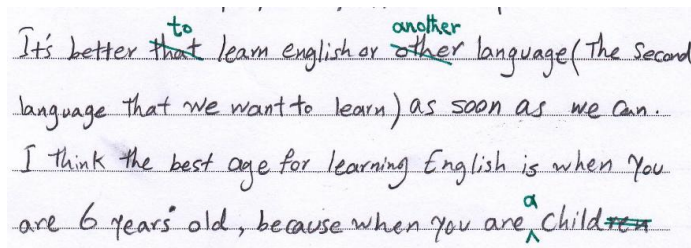
Excerpt 1. Sample feedback given to the WCF group

Group 2 received feedback on content and organization (COR), which addressed relevance, consistency in developing ideas, reasoning, sufficient support, and clarity as well as paragraphing, presence of topic sentences, grouping ideas, and conciseness. Comments incorporated criticism, suggestions, and occasionally praise written as imperatives, clarification requests, and (mitigated) interrogatives. Excerpt 2 shows a sample of feedback on content and organization.
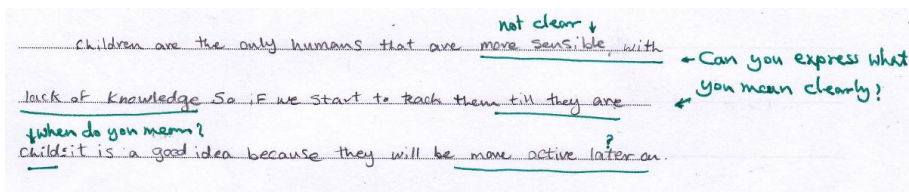
Excerpt 2. Sample feedback given the COR group

Group 3 received multilateral feedback (MUL), operationalized as a combination of written CF and feedback on content and organization. Excerpt 3 shows an example of MUL feedback.
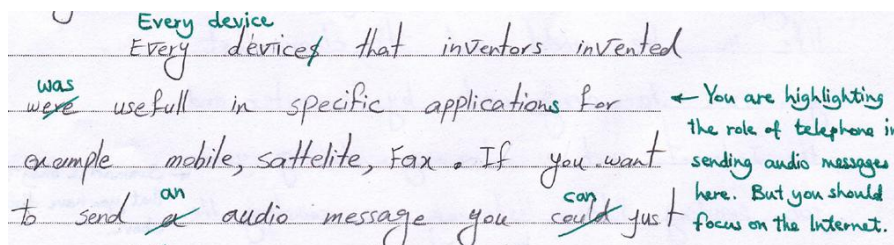
Excerpt 3. Sample feedback given to the MUL group

Group 4 received no feedback (NOF) until the study ended (when they received detailed feedback on five of their draft compositions). During the experiment, the participants were asked to rewrite their composition trying to improve it on their own, as shown in Excerpt 4.
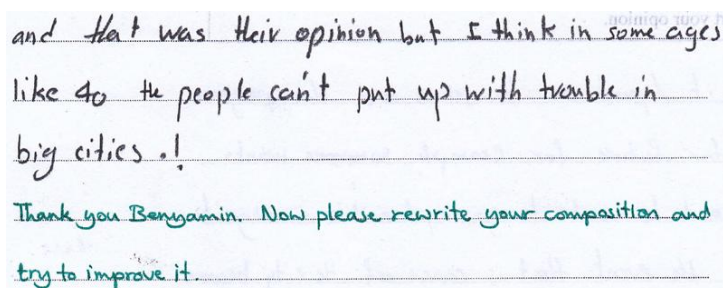
Excerpt 4. Reminder to the NOF group to rewrite

Procedures

In Week 1, following an introduction and ethics clearance, the pre-test was administered. The participants were allotted 45 minutes to write a composition on the given topic. The pre-test scripts were collected and were not returned to them. The eight classes were then randomly assigned to one of the four groups. From Week 2 to 9, during the writing sessions, the participants wrote eight different compositions. Each time, the scripts were collected, scanned and returned to the participants one or two days later with the relevant feedback (or no feedback in the case of the control group). Each treatment group received only the feedback which they were planned to receive. CF was given to the WCF and MUL groups. CF was mainly unfocused ranging from tenses, verb forms, voice, subject-verb agreement, word forms (e.g., success/succeed), plural vs. singular nouns, articles, prepositions, pronouns, possessive adjectives, and so forth. Errors in the use of more complex structures (e.g., relative pronouns), though, which had not been practiced were left uncorrected. Feedback on content and organization was given only to the MUL and COR groups.

The treatment groups were required to revise their compositions incorporating the comments or corrections, whereas the control group needed to revise their compositions making any changes which they believed would improve their first drafts. Each week, the participants submitted revisions prior to writing subsequent classroom compositions. In Week 10 the participants completed the post-test, without feedback, and in Week 14, the delayed post-test, both on the same topic as the pre-test.

Data coding procedure

In total, 92 students produced 19 scripts each (N = 1748): one at the pre-test (Script 1), one at the post-test (Script 10), one at the delayed post-test (Script 11), and eight interim compositions, which were also rewritten (Scripts 2 – 9). All test scripts at the pre-tests, post-tests, and delayed post-test (N = 276) (henceforth, Times 1, 2, 3) were assessed, with scores given to grammatical accuracy, complexity, fluency, content, and organization.

To measure accuracy, we calculated the number of errors per 100 words (EP100W), the number of errors divided by the total number of words multiplied by100 (see also Chandler, 2003 and Van Beuningen et al., 2012). We then coded the texts for clauses (see the definition in Polio, 1997). Following Wolfe-Quintero, Inagaki and Kim (1998), we calculated the percentage of error free clauses (EFC) by dividing the number of EFC by the total number of clauses multiplied by 100. To measure syntactic complexity, we used mean length of clause (Wolfe-Quintero et al., 1998), i.e., the number of words divided by the total number of clauses in each script. We measured fluency in terms of total words, as in Hartshorn et al.'s (2010) study.

To assess content and organization, two independent scales with four-level descriptors were developed. The scales were partially informed by the work of Jacobs, Zinkgraf, Wormuth, Hartfiel, and Hughey (1981); Brown and Bailey's (1984) scales; IELTS Task 2 writing band descriptors; and feedback from a panel of L2 testing experts. Content was operationalized with respect to five constructs: writer's position, relevance, reasoning, adequacy, and clarity. Organization was operationalized with respect to four constructs: structure of paragraphs, logical flow of ideas, cohesion, and paragraphing. A score between 1 and 4 was given to each of these constructs in content and organization scales on each script. The ultimate score ranged between 5 and 20 for content and between 4 and 16 for organization.

Inter-rater reliability and statistical analyses

Ten percent of all test scripts were randomly selected and double-coded by a trained coder to calculate inter-rater reliability. The Pearson Product Moment Correlation (r) was .99 for the number of clauses, 1.00 for the number of words, .92 for the number of errors, .97 for the number of EFC, .86 for the percentage of EFC, .83 for content, and .96 for organization.

In preparation for the statistical analyses, the equality of variance and normality of distribution were checked. Homogeneity of variances was verified through a non-significant Levene's test (Table 1). Kolmogorov-Smirnov and Shapiro-Wilk tests as well as visual inspection of histograms, normal Q-Q plots and boxplots showed that some variables were not normally distributed, and therefore *p*-values for the ANOVAs were obtained using non-parametric/distribution-free permutation tests with 4999 random permutations for all data (Efron & Tibshirani, 1998). The standard, parametric F-tests associated with ANOVAs assume that the 'errors' are independent and normally distributed with constant variance. For most variables, the permutation and parametric tests yielded very similar *p*-values.

Table 1
Levene's Test Showing the Homogeneity of Variances

Variable	Levene Statistic		df1	df2	Sig.
EFC	.413	3	88	.744	
EP100W		.343	3	88	.794
Syntactic Complexity	.093	3	88	.964	
Fluency		.399	3	88	.754
Content		.622	3	88	.603
Organization	.609	3	88	.611	

The data formed a three-by-four (time x group) factorial design. Comparisons of relevant pairs of means were made based on Tukey's post-hoc tests. In what follows below, the Tukey's tests are reported as this procedure corrects for multiple comparisons. In addition, Cohen's *d* was computed to estimate effect sizes. Cohen's *d* represents the mean difference standardized by standard deviations. Following Cohen (1988), the effect sizes were interpreted as follows: small = .2, medium = .5, and large = .8.

Results

In the following sections, the descriptive and inferential statistics of the impact of the three feedback types and of no feedback on the dependent variables of the study are presented.

Impact of feedback on accuracy

Table 2 displays the descriptive statistics including mean scores and standard deviations of accuracy measured by the percentage of EFC and number of EP100W over the three main writing occasions, Time 1, Time 2 and Time 3 for the four groups receiving, (1) written corrective feedback (WCF), (2) feedback on content and organization (COR), (3) multilateral feedback (MUL), and (4) no feedback (NOF).

Table 2
Descriptive Statistics for the Two Accuracy Measures

Group	N	Accuracy (Percentage of EFC)						Accuracy (Number of EP100W)					
		Pre-test		Post-test		Delayed post-test		Pre-test		Post-test		Delayed post-test	
		M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
1. WCF.	25	35.25	16.05	52.27	19.51	48.16	17.88	13.51	4.79	9.12	4.15	10.18	5.08
2. COR.	19	37.79	13.80	35.12	16.07	33.50	18.64	14.36	6.86	13.90	5.40	12.50	5.05
3. MUL.	26	38.73	13.93	45.87	16.94	49.12	15.49	13.97	6.66	9.92	4.70	9.86	5.76
4. NOF.	22	39.00	15.55	39.44	16.72	41.05	15.82	14.10	6.54	13.46	5.41	13.12	6.66

A series of one-way ANOVA tests with groups as a between-subject variable performed with regard to the two accuracy measures showed no statistically significant differences amongst the groups at the time of the pre-test when the percentage of EFC was considered [$F(3,88) = 0.322$, $p = 0.809$], and when the number of EP100W was considered [$F(3,88) = 0.075$, $p = 0.973$]. To establish whether there were significant differences in accuracy levels across the feedback groups over time, a series of two-way repeated measures ANOVA were performed with *p*-values obtained from the permutation tests (Table 3). As can be seen, there were significant interaction

effects between Time and Group when accuracy was measured by the percentage of EFC (p <.001) and by the number of EP100W (p = .006), indicating that the four groups developed differentially over time.

Table 3
Repeated Measures ANOVA of Accuracy Analysis across Groups over Time

Source	df	F	p	Partial Eta Sq.
Percentage of EFC				
Between subjects				
Group	3	2.112	.035	.067
Within subject				
Time	2	9.756	<.001	.100
Time x Group	6	6.217	<.001	.175
Number of EP100W				
Between subjects				
Group	3	46.931	.090	
Within subject				
Time	2	28.815	<.001	.247
Time x Group	6	4.398	.006	.130

More detailed analyses of the Tukey's post-hoc test showed that the WCF group improved significantly in accuracy measured by the percentage of EFC with a large or medium effect size from Time 1 to both Time 2 (Cohen's d = 0.95) and Time 3 (Cohen's d = 0.76). Likewise, with regards to the number of EP100W, the WCF group improved significantly from Time 1 to both Time 2 (Cohen's d = 0.98) and Time 3 (Cohen's d = 0.67). Similarly, accuracy measured by the percentage of EFC in the MUL group improved significantly at Times 2 (Cohen's d = 0.46) and 3 (Cohen's d = 0.70). Moreover, the number of EP100W dropped significantly at Time 2 (Cohen's d = 0.70) and Time 3 (Cohen's d = 0.66).

By contrast, the percentage of EFC in the COR group continued to decline each time although neither decline was significant. Even though the participants in the COR group deteriorated in the percentage of EFC, they made a lower number of EP100W over Times 2 and 3. While this gain in accuracy was not significant at Time 2, it proved significant at Time 3 (Cohen's d = 0.31). Finally, in the NOF group, the accuracy levels measured by the percentage of EFC did not significantly grow from Time 1 to Times 2 and 3. The percentage of EP100W declined, though not significantly, from Time 1 to both Times 2 and 3.

Impact of feedback on syntactic complexity

Table 4 illustrates the descriptive statistics including mean scores and standard deviations of syntactic complexity in the four groups.

Table 4
Descriptive Statistics for Syntactic Complexity

		Syntactic complexity (Average number of words per clause)						
Group	N	Pre-test				Post-test		Delayed post-test
		M	SD	M	SD	M	SD	
1. WCF	25	8.09	1.50			8.15	1.81	8.68 1.88
2. COR	19	8.50	1.30			9.84	1.95	10.00 1.49
3. MUL	26	8.18	1.55			9.37	1.40	8.85 1.30
4. NOF	22	8.04	1.25			8.20	1.56	8.27 1.13

A series of one-way ANOVA with groups as a between-subject variable showed no significant differences at the time of the pre-test with respect to syntactic complexity [$F(3,88) = 0.420$, $p = 0.739$]. To examine whether the variations were statistically significant across the four groups' scores over time, a series of two-way repeated measures ANOVA were performed with clause length as the dependent variable and feedback groups as the independent variable. As Table 5 shows, the interaction between Time and Group was not significant with respect to syntactic complexity.

Table 5
Repeated Measures ANOVA of Syntactic Complexity across Groups over Time

Source	df	F	p	Partial Eta Sq.
Syntactic complexity				
Between subjects				
Group	3	4.685	.001	.138
Within subject				
Time	2	11.897	<.001	.119
Time x Group	6	3.465	.091	

Impact of feedback on written fluency

Table 6 illustrates the descriptive statistics including mean scores and standard deviations of fluency.

Table 6
Descriptive Statistics for Fluency

Group	N	Fluency (Average number of words per script)					
		Pre-test		Post-test.		Delayed post-test	
		M	SD	M	SD	M	SD
1. WCF	25	216.28	72.95	218.04	61.92	197.12	48.80
2. COR	19	206.21	60.76	252.21	44.69	236.84	58.19
3. MUL	26	208.96	44.87	232.96	42.25	219.04	48.18
4. NOF	22	216.00	46.19	199.86	42.57	193.86	43.34

A series of one-way ANOVA showed no statistically significant differences amongst the four groups at the time of the pre-test with respect to fluency [$F(3,88) = .171$, $p = .916$]. A series of repeated measures ANOVA showed that there were significant interactions between Time and Group ($p < .01$) with respect to fluency, indicating that the four groups developed differentially over time (Table 7).

To examine the differences precisely, post hoc comparisons were made using Tukey's post-hoc tests. The analyses revealed that fluency grew significantly in the COR group (Cohen's $d = 0.86$) and MUL group (Cohen's $d = 0.55$) from Time 1 to 2. While this growth was still significant for the COR group at Time 3 (Cohen's $d = 0.51$), it was not the case for the MUL group. On the other hand, the WCF group showed no significant fluctuation. The NOF group's fluency showed no significant change between Time 1 and 2, but that there was a significant decline from Time 1 to Time 3 (Cohen's $d = 0.49$).

Table 7
Repeated Measures ANOVA of Fluency across Groups over Time

Source	df	F	p	Partial Eta Sq.
Fluency				
Between subjects				
Group	3	1.625	.200	
Within subject				
Time	2	5.230	<.001	.056
Time x Group	6	4.359	.004	.129

Impact of feedback on content

Table 8 reflects the descriptive statistics including mean scores and standard deviations of content.

Table 8
Descriptive Statistics for Content

Group	N		Pre-test		Post-test		Delayed post-test	
		M	SD	M	SD	M	SD	
1. WCF	25	10.68	2.04	10.56	2.31	10.80	2.57	
2. COR	19	11.11	2.47	13.84	2.39	13.11	3.21	
3. MUL	26	11.69	2.49	13.08	2.87	12.31	2.54	
4. NOF	22	11.14	2.30	10.73	2.66	10.14	1.98	

A one-way ANOVA showed no statistically significant difference between the four groups [$F(3,88) = .814$, $p = .489$] in pre-test scores with respect to content. A series of repeated measures ANOVA showed that there were significant interactions between Time and Group ($p < .001$) with respect to content, indicating that the four groups developed differentially over time (Table 9).

Table 9
Repeated Measures ANOVA of Content across Groups over Time

Source	df	F	p	Partial Eta Sq.
Content				
Between subjects				
Group	3	5.899	.001	.167
Within subject				
Time	2	6.545	.002	.069
Time x Group	6	4.870	.000	.142

Tukey's post-hoc analyses revealed that the performance measured by content grew significantly in the COR group (Cohen's $d = 1.12$) and MUL group (Cohen's $d = 0.52$) from Time 1 to 2. While this growth was still significant for the COR group at Time 3 (Cohen's $d = 0.70$), it was not the case for the MUL group. On the other hand, fluctuations by the WCF and NOF group at Times 2 and 3 were not significant.

Impact of feedback on organization

Table 10 shows the descriptive statistics including mean scores and standard deviations of the four groups.

Table 10
Descriptive statistics for organization

Group	N	Organization (structure of units of idea, logical flow, cohesion, and paragraphing)						
		Pre-test		Post-test		Delayed post-test		
		M	SD	M	SD	M	SD	
1. WCF	25	7.00	2.10	7.20	1.53	7.16	2.06	
2. COR	19	7.84	2.36	11.74	2.86	12.26	2.31	
3. MUL	26	7.54	2.39	12.19	2.12	11.73	2.05	
4. NOF	22	7.05	1.81	7.27	1.24	7.09	1.54	

A one-way ANOVA showed no statistically significant difference between the four groups [$F(3,88) = .745$, $p = .528$] at the time of the pre-test in terms of organization. Repeated measures ANOVA showed that there were significant interactions between Time and Group ($p < .001$) with respect to organization (Table 11).

Table 11
Repeated Measures ANOVA of Organization across Groups over Time

Source	df	F	p	Partial Eta Sq.
Organization				
Between subjects				
Group	3	30.679	.000	.511
Within subject				
Time	2	75.323	.000	.461
Time x Group	6	22.689	.000	.436

The Tukey's post-hoc test showed that the performance measured by organization grew significantly, with very large effect sizes, in the COR group (Cohen's $d = 1.49$) and MUL group (Cohen's $d = 2.06$) from Time 1 to 2. The COR group retained its significant level at Time 3 (Cohen's $d = 1.89$), and so did the MUL group (Cohen's $d = 1.88$). In contrast, there was little change in performance made by the WCF and NOF groups.

In a nutshell, providing WCF significantly improved grammatical accuracy, but not fluency, content and organization. On the other hand, only those writers who received feedback on content and organization improved in these two aspects as well as in fluency at least in the post-test.

Discussion

Accuracy gain through CF

The first research question explored whether sustained teacher feedback on grammar, content and organization led to improved accuracy. The findings revealed that only the groups who received written CF improved significantly in accuracy. These findings corroborate those in unfocused

written CF studies (e.g., Boggs, 2019; Bonilla-Lopez, Steendam, Speelman & Buyse, 2018; Chandler, 2003; Ellis et al., 2008; Hartshorn et al., 2010; Van Beuningen et al., 2012). What distinguishes our study is that all groups engaged in the same number of (re)writing activities, and the results were compared against those of a no-feedback group. The finding contradicts Sheen et al.'s (2009) explanation that when the volume of CF extends beyond a limited scope, feedback recipients cannot process CF. Importantly, accuracy improved in the group who received multilateral feedback, which incorporated CF, without leading to cognitive overload with respect to attention to grammatical errors. This finding, therefore, contradicts Ellis et al.'s (2008) assumption that "[a] mass of corrections directed at a diverse set of linguistic phenomena (and perhaps also at content and organisational issues) is hardly likely to foster the noticing and cognizing that may be needed for CF to work for acquisition" (p. 368).

A comparison of the two measures of accuracy show that the MUL group reduced the number of errors and composed more EFCs in the post-test, with a medium and small effect size, respectively, whereas the WCF group did so with a large effect size. This can be attributed to sheer focus of feedback on grammar in compositions written by the latter group. The delayed post-test findings suggest that the CF recipients were able to retrieve and employ the knowledge they had consolidated to write more accurately one month later. Bitchener (2012) argues that retaining improvement over time verifies "the potential for learners to retrieve the explicit knowledge gained from written CF and stored in their long-term memory" (p. 353). These key findings provide further evidence for the internalization of knowledge in the students' long-term memory which we suggest was due to the combined attention to CF (Schmidt, 2001) and prolonged engagement in writing (DeKeyser, 1997).

Feedback exclusively targeting ideational and organizational issues did not lead to improved grammatical accuracy. We hypothesize that students did not attend to grammatical errors when not encouraged to do so by the feedback provided. Furthermore, providing no CF and mere engagement in sustained (re)writing activities, as in the case of the control group in our study, did not lead to improved accuracy. The findings of the no-feedback group negate Truscott's (1996) argument for adequacy of student writing practice in improving written accuracy. The findings seem to reinforce Bitchener and Ferris's (2012) standpoint that self-editing requires some intervention, such as teacher feedback, and Bruton's (2009b, p. 604) argument for the necessity of teacher feedback: "there can be no new language development from just writing. Language development has to be prompted by some kind of further input, corrective or otherwise."

Theoretically, the current study endorses the importance of awareness at levels of noticing and understanding underlined in the noticing hypothesis (Schmidt, 2001). Sustained CF enabled the CF recipients to not only notice feedback on numerous grammatical categories, but also retain their knowledge and extrapolate the rules governing those structures when writing new compositions even without complementary metalinguistic explanation. The findings help us hypothesize that prolonged practice (DeKeyser, 1997) is likely to result in accuracy gain as long as it is guided by written CF. It is conceivable that when the scope of target structures extends, CF recipients need a larger number of treatments to engage with CF to notice their various errors and compose more accurate texts. We cannot however argue about the influence of sustained feedback on the development of implicit knowledge and proceduralization (Ellis, 2011). We are not confident whether part of the explicit knowledge students gained through sustained CF converted into implicit knowledge (Storch, 2010). Rather we argue that the L2 writers expanded their explicit knowledge of written accuracy by drawing on the CF which they continued to receive, thereby consciously monitoring what they were generating in subsequent writing given the time they had during writing (Polio, 2012). Another explanation might be that the participants had already attained metalinguistic knowledge of grammar (e.g., tense and verb form) through instruction elsewhere (e.g., at school), and the CF only drew their attention to overall accuracy [See Shintani et al. (2014) for further discussion]. Importantly, that CF recipients retained their

accuracy knowledge suggests that they were able to store part of the explicit knowledge that they attained in their long-term memory and retrieve the knowledge after a month. Of note is the ability of the recipients of multilateral feedback in this respect.

Gains in Syntactic Complexity and Fluency

The findings showed that sustained attention to grammatical errors improved written accuracy, without detriment to syntactic complexity and fluency. The post-test findings related to the WCF group concur with other written CF studies (e.g., Chandler, 2003; Hartshorn et al., 2010; Soleimani, et al. 2017; Van Beuningen et al., 2012). In Soleimani, et al. (2017), the advanced and intermediate students who wrote collaboratively even outperformed the individual writers in fluency measured by average number of words, clauses and T-units across drafts of the same composition. In our study, however, we observed a fall (though not significant) in fluency across compositions. The observed fall in fluency for the WCF group may be attributable to a trade-off between accuracy and fluency (Skehan, 1998). Moreover, the findings showed that under similar circumstances an increase in composition length is an artefact of student response to feedback on content and organization. In fact, it is feasible to offer CF to compositions and, unlike the debated deterioration, still anticipate improvement with medium or large effect size in fluency, providing feedback incorporates comments on content and organization.

Gains in Content and Organization

Research question 3 explored the influence of sustained feedback on content. The findings demonstrated that only the recipients of sustained feedback on content and organization showed significant improvement in content quality in the post-test with medium to large effect sizes. The findings corroborate those in Kepner's (1991) and Paulus's (1999) studies, which suggest that CF alone or no feedback is unlikely to improve the quality of L2 writing. It is likely that writing on various topics enables feedback recipients to explore and practice content-governing conventions, even when feedback also addresses grammatical accuracy and organization. The findings also revealed that sustained feedback on organization – along with feedback on content or altogether with CF – is conducive to learning organizational conventions of a composition. Sustained practice alone did not lead to improved organization. For example, many non-recipients of such feedback still omitted a clear conclusion or introduction. We could not find other L2 written feedback studies, where organization was independently and objectively assessed. Some (e.g., Hartshorn et al., 2010) assessed it as a construct subsumed under rhetorical competence. They found no significant difference in written content between their recipients and non-recipients of rhetorical feedback.

Balanced Development in L2 Writing

Research question 5 explored the extent of multidimensional writing development through sustained multilateral feedback. It appears that providing L2 writers with sustained feedback on accuracy, content and organization can result in balanced improvement in all aspects of writing. Sustained guided writing can enable L2 writers to not only "learn to write" (Hyland, 2011) to convey their message across to an audience, but also "write to learn" (Manchón, 2011a), i.e., to consolidate and develop knowledge in L2. "The wider the evidence-base, the more certain we can be that a learner has acquired features of the L2" (Bitchener, 2012, p. 349). Pedagogically, this finding confirms teachers' practices of adopting a multilateral approach in L2 writing classes (Hyland, 2003).

With respect to the findings of multilateral feedback (regardless of instruction), Figure 1 shows a procedure which can contribute to L2 writing development. The procedure starts with the L2 writer composing an exposition, followed by the teacher's feedback on aspects of language, content, and organization, the writer's gap-noticing, and redrafting the composition guided by feedback prior to writing a subsequent composition. The repetition of arrows signifies the recurrence of this cycle.
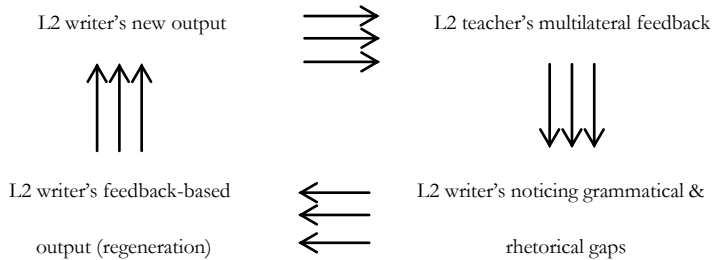

Figure 1. Recurrence of writing, providing feedback and rewriting

Conclusion

The participants in this study were EFL learners of low- to mid-intermediate proficiency who were asked to participate in a weekly writing session in addition to their daily conversation courses. It can be argued that voluntary participants are often highly motivated. This suggests that our findings may not be generalizable to classroom contexts where learners may not be as highly motivated. The findings of this study endorse the language learning potentials of writing (Manchón, 2011b) suggesting that sustained written output practice guided by teacher feedback, on which L2 writers consistently act, contributes to L2 writing development. One principal pedagogical implication of this study is that L2 teachers should continue to provide written feedback on those dimensions of writing which need improvement. Because most high-stakes testing systems, practitioners, and L2 learners greatly value accuracy gain, the current study recommends the provision of sustained multilateral written feedback, which target rhetorical feedback and feedback on linguistic features.

The study has other pedagogical (and empirical) implications. First, concerning the scope of target structures, the study introduced a relatively unfocused approach with an extended number of targeted structures. This is a compromise between the focused approach to CF, which is pedagogically less viable (Bruton, 2009b; Storch, 2010; Xu, 2009) and the unfocused approach, which might be overwhelming to CF recipients (Bitchener, 2012; Ferris, 2010), especially those of lower proficiency levels. Moreover, concerning the ethical issue of engagement of the control group in sustained (re)writing, feedback was given to this group at the end of the study. Although delayed feedback is never as motivating for students as timely feedback (Hartshorn et al., 2010), this can be a good compromise to the ethical issue of not providing feedback to learners. We suggest that further studies be undertaken to explore the influence of sustained written feedback on content, organization and CAF in other settings. Because providing feedback on different aspects of L2 writing is often practiced by L2 writing practitioners, future research in this area will shed light on whether or not this teacher practice is beneficial. Based on the findings of this study we suggest that L2 teachers continue to offer sustained written feedback on L2 compositions.

References

Ashwell, T. (2000). Patterns 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*(3), 227–258. DOI: 10.1016/S1060-3743(00)00027-8.

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. (2012). A reflection on 'the language learning potential' of written CF. *Journal of Second Language Writing, 21*(4), 348-363. DOI: 10.1016/j.jslw.2012.09.006.

Bitchener, J., & Ferris, D. R. (2012). *Written corrective feedback in second language acquisition and writing.* New York: Routledge. DOI: 10.1111/j.1540-4781.2013.12000.x.

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

Boggs, A. J. (2019). Effects of teacher-scaffolded and self-scaffolded corrective feedback compared to direct corrective feedback on grammatical accuracy in English L2 writing. *Journal of Second Language Writing, 46*(4) 100671. DOI: 10.1016/j.jslw.2019.100671

Bonilla López, M., Steendam, E. V., Speelman, D. & Buyse, K. (2018). The differential effects of comprehensive feedback forms in the second language writing class. *Language Learning, 68*(3). 813-850. DOI: 10.1111/lang.12295.

Brown, J. D., & Bailey, K. M. (1984). A categorical instrument for scoring second language writing skills. *Language Learning, 34*, 21-42. DOI: 10.1111/j.1467-1770.1984.tb00350.x.

Bruton, A. (2009a). Designing research into the effect of error correction in L2 writing: Not so straightforward. *Journal of Second Language Writing, 18*(2), 136-140. DOI: 10.1016/j.jslw.2009.02.005.

Bruton, A. (2009b). Improving accuracy is not the only reason for writing, and even if it were.... *System 37*(4), 600-613. DOI: 10.1016/j.system.2009.09.005.

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. DOI: 10.1016/S1060-3743(03)00038-9.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates. DOI: 10.1016/B978-0-12-179060-8.50012-8.

DeKeyser, R. (1997). Beyond explicit rule learning: Automatizing second language morphosyntax. *Studies in Second Language Acquisition 19*(2), 195-221.

DeKeyser, R. (2007). Skill acquisition theory. In B. VanPatten, & J. Williams (Eds.), *Theories in Second Language Acquisition* (pp. 97-113). Lawrence Earlbaum Associates, Publishers.

Efron, B., & Tibshirani, R. (1998). The problem of regions. *Annals of Statistics, 26*(5), 1687-1718.

Ellis, N. (1994). *Implicit and explicit learning of languages*. London: Academic Press.

Ellis, N. (2011). Implicit and explicit SLA and their interface. In C. Sanz & R. Leow (Eds.), *Implicit and explicit language learning: Conditions, processes, and knowledge in SLA and Bilingualism* (pp. 35-47). Washington, DC: Georgetown University Press. DOI: 10.1111/j.1540-4781.2012.1311.x.

Ellis, R. (1994). *The study of second language acquisition*. Oxford: Oxford University Press.

Ellis, R. (2009). Implicit and explicit learning, knowledge and instruction. In R. Ellis, S. Loewen, C. Elder, H. Reinders, R. Erlam, & J.Philp (Eds.),*Implicit and Explicit Knowledge in Second Language Learning, Testing and Teaching* (pp. 3-25). Multilingual Matters. DOI: 10.1111/j.1540-4781.2011.01253.x.

Ellis, R., Sheen, Y., Murakami, M., & Takashima, H. (2008). The effects of focused and unfocused written corrective feedback in English as a foreign language context. *System, 36*(3), 353-371. DOI: 10.1016/j.system.2008.02.001.

Ferris, D. (1997). The influence of teacher commentary on student revision. *TESOL Quarterly, 31*(2), 315-339. DOI: 10.2307/3588049.

Ferris, D. (1999). The case for grammar correction in L2 writing classes: A response to Truscott (1996). *Journal of Second Language Writing, 8*(1), 1-10. DOI: 10.1016/S1060-3743(99)80110-6.

Ferris, D. (2006). Does error feedback help student writers? New evidence on the short- and long-term effects of written error correction. In K. Hyland & F. Hyland (Eds.), *Feedback in second language writing: Contexts and issues* (pp. 81-104). Cambridge: Cambridge University Press. DOI: 10.1017/S0261444806003399.

Ferris, D. (2010). Second language writing research and written corrective feedback in SLA. *Studies in Second Language Acquisition, 32*(2), 181-201. DOI: https://doi.org/10.1017/ S0272263109990490.

Frear, D. & Chiu, Y. (2015). The effect of focused and unfocused indirect written corrective feedback on EFL learners' accuracy in new pieces of writing. *System, 53*, 24-34. DOI: 10.1016/j.system.2015.06.006.

Gass, S. (2003). Input and interaction. In Doughty, C & Long, M. (Eds.), *The Handbook of Second Language Acquisition* (pp. 224-255). Oxford, England: Blackwell.

Hartshorn, K. J., Evans, N. W., Merrill, P. F., Sudweeks, R. R., Strong-Krause, D., & Anderson, N. J. (2010). Effects of dynamic corrective feedback on ESL writing accuracy. *TESOL Quarterly, 44*(1), 84-109. DOI: 10.5054/tq.2010.213781.

Hyland, F. (2003). Focusing on form: Student engagement with teacher feedback. *System*, *31*(2), 217-230. DOI: 10.1016/S0346-251X(03)00021-6.

Hyland, F. (2011). The language learning potential of form-focused feedback on writing: students' and teachers' perceptions. In R. M. Manchón (Ed.), *Learning to write and writing to learn in an additional language* (pp. 159-179).Amsterdam: John Benjamins. DOI: 10.1075/lllt.31.

Jacobs, H. L., Zinkgraf, S. A., Wormuth, D. R., Hartfiel, V. F., & Hughey, J. B. (1981). *Testing English Composition: A Practical Approach*. Rowley, Mass.: Newbury House.

Kang, E., & Han, Z. (2015). The Efficacy of written corrective feedback in improving L2 written accuracy: a meta-analysis. *Modern Language Journal. 99*(1), 1–18. DOI: 10.1111/modl.12189.

Kao, C.-W., & Wible, D. (2014). A meta-analysis on the effectiveness of grammar correction in second language writing. *English Teaching & Learning, 38*(3), 29–69. DOI: 10.6330/ETL.2014.38.3.02.

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. DOI:10.1111/j.1540-4781.1991.tb05359.x.

Lee, I. (2019). Teacher written corrective feedback: Less is more. *Language Teaching, 52*(4), 524-536. DOI:10.1017/S0261444819000247

Manchón, R. M. (2011a). The language learning potential of writing in foreign language contexts: Lessons from research. In T. Cimasko & M. Reichelt (Eds.), *Foreign language writing instruction: Principles and practices* (pp. 44-64). Anderson, SC: Parlor Press. DOI: 10.1558/wap.v5i1.151.

Manchón, R. M. (2011b). Writing to learn the language: Issues in theory and research. In R. M. Manchon (Ed.), *Learning-to-write and writing-to-learn in an additional language* (pp. 61-82). Amsterdam: John Benjamins. DOI: 10.1075/lllt.31.

Manchón, R. M., & Roca de Larios. J. (2011). Writing to learn in FL contexts: Exploring learners' perceptions of the learning potential of L2 writing. In Manchón R. M. (Ed.) *Learning-to-write and writing-to-learn in an additional language* (pp. 181-207). Amsterdam: John Benjamins. DOI: 10.1075/lllt.31.

Pashazadeh, A. (2017). The Effect of Mid-Focused and Unfocused Written Corrections on the Acquisition of Grammatical Structures. *Journal of Response to Writing, 3*(1), 56-82.

Paulus, T. M. (1999). The effect of peer and teacher feedback on student writing. *Journal of Second Language Writing, 8*(3), 265-289. DOI: 10.1016/S1060-3743(99)80117-9.

Polio, C. (1997). Measures of linguistic accuracy in second language writing research. *Language Learning, 47*(1), 101-143. DOI: 10.1111/0023-8333.31997003.

Polio, C. (2012). The relevance of second language acquisition theory to the written error correction debate. *Journal of Second Language Writing, 21*, 375-389. DOI: 10.1016/j.jslw.2012.09.004.

Riazantseva, A. (2012). Outcome measure of L2 writing as a mediator of the effects of corrective feedback on students' ability to write accurately. *System, 40*(3), 421-430. DOI: 10.1016/j.system.2012.07.005.

Rummel, S., & Bitchener, J. (2015). The effectiveness of written corrective feedback and the impact Lao learners' beliefs have on uptake. *Australian Review of Applied Linguistics, 38*(1), 64-82. DOI: 10.1075/aral.38.1.04rum.

Russell, J., & Spada, N. (2006). The effectiveness of corrective feedback for the acquisition of L2 grammar: A meta-analysis of the research. In J. M. Norris & L. Ortega (Eds.), *Synthesizing Research on Language Learning and Teaching* (pp. 133-164).Amsterdam: John Benjamins Publishing Company.

Schmidt, R. (1990). The Role of Consciousness in Second Language Learning. *Applied Linguistics, 11*(2), 129-158. DOI: 10.1093/applin/11.2.129.

Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and Second Language Instruction* (pp. 3-32). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139524780.

Semke, H. (1984). The effects of the red pen. *Foreign Language Annuals, 17*, 195-202. DOI: 10.1111/j.1944-9720.1984.tb01727.x.

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.x.

Sheen, Y., Wright, D., &Moldawa, A. (2009). Differential effects of focused and unfocused written correction on the accurate use of grammatical forms by adult ESL learners. *System, 37*(2), 556-569. DOI: 10.1016/j.system.2009.09.002.

Sheppard, K. (1992). Two feedback types: Do they make a difference? *RELC Journal, 23*, 103-110. DOI: 10.1177/003368829202300107.

Shintani, N., & Ellis, R. (2013). The comparative effect of direct written corrective feedback and metalinguistic explanation on learners' explicit and implicit knowledge of the English indefinite article. *Journal of Second Language Writing, 22*(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.

Skehan, P. (1998). *A Cognitive Approach to Language Learning.* Oxford: Oxford University Press.

Storch, N. (2010). Critical feedback on written corrective feedback research. *International Journal of English Studies, 10*(2), 29-46.

Suzuki, W., Nassaji, H., & Sato, K. (2019). The effects of feedback explicitness and type of target structure on accuracy in revision and new pieces of writing. *System, 81*, 135-145. DOI: 10.1016/j.system.2018.12.017

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.tb01238.x.

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.2008.05.003.

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

Wolfe-Quintero, K., Inagaki, S., & Kim, H. Y. (1998). *Second Language Development in Writing: Measures of Fluency, Accuracy, and Syntactic Complexity*. Honolulu: University of Hawaii.

Xu, C. (2009). Overgeneralization from a narrow focus: A response to Ellis et al. (2008) and Bitchener (2008). *Journal of Second Language Writing*, *18*(4), 270-275. DOI: 10.1016/j.jslw.2009.05.005.

Acknowledgment

Ali Rastgou holds PhD in Applied and Educational Linguistics from the University of Melbourne, and is the Head of Linguistics and TEFL Department, Islamic Azad University, Shahrood Branch. He has presented at AAAL, SSLW and ALTAANZ, and is the Academic Editor of high school books, Vision 2 & 3. His research focuses on L2 writing development.

Neomy Storch is an Associate Professor in ESL and Applied Linguistics at the School of Languages & Linguistics, the University of Melbourne. Her research focuses on topics related to L2 writing, including collaborative writing, corrective feedback, writing development and authorial voice. She has published and presented extensively on her research.

Ute Knoch is an Associate Professor and the Director of the Language Testing Research Centre at the University of Melbourne. Her research interests are in the area of writing assessment (including effective feedback provision), rating processes, test validation, assessing languages for academic and professional purposes, and placement testing.