



A Meta-Analysis of Short- and Long-Term Effects of Written Corrective Feedback on L2 Grammatical Accuracy in Writing

EunYoung Kang (Kongju National University)



This is an open-access article distributed under the terms of the Creative Commons License, which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Received: November 06, 2022

Revised: November 17, 2022

Accepted: November 30, 2022

Eun Young Kang
Assistant Professor, Division of
Liberal Arts, Kongju National
University
Tel: (041)-521-9735
E-mail: ekang@kongju.ac.kr

ABSTRACT

Kang, EunYoung, 2022. A meta-analysis of short- and long-term effects of written corrective feedback. *Korean Journal of English Language and Linguistics* 22, 1269-1286.

Written Corrective Feedback (WCF) is one of the most widely implemented teaching strategies in the second language (L2) writing classes. While the evidence about the positive effects of WCF on L2 grammatical accuracy has been reported, comparatively little is known about its long-term effects and related factors that moderate the effects. This meta-analysis aims to report and compare the short- and long-term effects of WCF on improving L2 grammatical accuracy. The data set for this study involved 25 primary studies exploring both short- and long-term effects of WCF. The study found that WCF yielded a positive effect on L2 accuracy, compared to the no-WCF condition. The overall effect of WCF was moderate ($g = 0.62$) on immediate posttests but fell within the small-to-moderate range ($g = 0.46$) on delayed posttests. As for moderator variables, they functioned in a similar fashion depending on the two test time points. Specifically, the genre of writing tasks and types of feedback were the significant mitigating factors on the immediate and delayed posttests.

KEYWORDS

written corrective feedback, long-term effects, meta-analysis, L2 accuracy in writing

1. Introduction

In the second language acquisition (SLA) literature, written corrective feedback (WCF) indicates any correction the instructor offers on learners' writing errors. Feedback can be provided on ideas or the organization of writing, but SLA researchers have been mainly interested in the effects of WCF on learners' grammatical accuracy in writing (Ferris 2010). Various types of WCF have been investigated, ranging from indirect (only indicating the existence of an error) to direct (providing the correct form of each error) and from unfocused (correcting every error) to focused feedback (correcting targeted linguistic errors).

Previous studies have suggested that WCF can be an effective tool to improve second language (L2) learners' accuracy in writing (e.g., Ashwell 2000, Fathman and Whalley 1990, Ferris and Roberts 2001, Kang and Han 2015). Although the positive effects of WCF in the short term have been well-documented, many empirical studies have not reported any results with regard to the long-term effects of WCF. More recently, however, an increasing number of researchers have investigated how learners' accuracy in writing improves over the long term in response to WCF. It is of obviously great interest to both researchers and teachers to transcend the study of short-term effects to comprehend whether the provision of WCF leads to longer-term gains in grammatical accuracy and which features of WCF relate to learning outcomes. Therefore, the present meta-analysis focuses on WCF studies employing both immediate and delayed posttests to evaluate and compare the short- and long-term effects of WCF directly. In addition to examining the short- and long-term effects of WCF with a sample of recent studies, moderator variables were examined to shed light on reasons for effect size changes from immediate posttests to delayed posttests.

2. Literature Review

Since the mid-90s, interest in WCF research has expanded, spurred mainly by Truscott's (1996) critical review of extant research on error correction and strong argument against WCF. Truscott argued poignantly that WCF should be abandoned entirely. He asserted that error correction is futile because corrections are not always in tune with the learner's level of grammatical knowledge. He noted that the structures targeted in a teacher's WCF might not necessarily be the ones the student is developmentally ready to acquire. He further claimed that WCF leads only to "a superficial and possibly transient form of knowledge" (p. 345) and does not help develop L2 writers' long-term accuracy.

In response to these arguments, much empirical research on the efficacy of WCF was conducted in the following years. Most of the studies provided counter-evidence to Truscott's claims, indicating that WCF can improve L2 accuracy (see, e.g., Ashwell 2000, Chandler 2003, Ferris 2006). While early research on WCF examined whether WCF promotes L2 learners' accuracy in writing in general, ensuing research went beyond this issue, exploring finer-grained problems such as what type of feedback is more effective.

As for the type of feedback, a majority of studies have attempted to investigate either direct or indirect WCF. Direct WCF provides the correct form of an error, while indirect WCF simply indicates the presence of errors by underlining, circling, or using special symbols to hint at the nature of the errors, such as articles and subject-verb agreement (Ellis 2008). The main distinction of indirect WCF from direct feedback is that learners are expected to self-correct their errors based on these cues. Researchers have proposed divergent views with regard to the relative effectiveness of these two types of WCF. Direct feedback has been found to be more effective in reducing students' errors as it is more immediate and specifically tackles errors in students' writing (Chandler 2003, Ferris

and Hedgecock 2014, Ferris and Roberts 2001). However, other researchers have found the opposite, noting that indirect WCF is more beneficial because of learners' engagement and reflection on their own errors in writing. Still, others have revealed that indirect feedback only works if learners have sufficient grammatical knowledge to self-correct the errors in question (Ellis et al. 2008).

The scope of WCF (focused or unfocused) has also received some scholarly attention in the WCF literature. Focused WCF entails selective corrections of one or a few specific error types, whereas unfocused feedback involves comprehensive corrections of every error type (Ellis 2008). Focusing on one or two types of preselected linguistic errors could ease L2 learners' cognitive burden and draw their attention to the errors, which in turn may help them learn these forms (Bitchener 2008, Bitchener and Knoch 2010, Sheen 2007). However, it may not be a suitable corrective strategy if the goal of instruction is to improve students' overall written accuracy (Karim and Nassaji 2018).

As noted above, various claims have been made in terms of the relative effectiveness of feedback strategies. Several researchers have attempted to resolve discrepancies across individual studies by synthesizing empirical studies on WCF via meta-analysis. Truscott's (2007) study is arguably the first meta-analysis on L2 WCF. He reported a small effect size ($d = 0.148$) for six studies with a pre- and a posttest design and -0.155 for six studies with a treatment/control design, concluding that the effects of WCF on L2 accuracy are minimal. However, the results may not be generalizable since only a small number of studies were included. He also did not investigate how moderator variables related to WCF play a part in the effectiveness of WCF.

To address these issues, Kang and Han (2015) took a meta-analytic approach to synthesize 22 empirical studies on WCF, including a control group. The rationale behind only including controlled studies was to present a more valid investigation of the effects of WCF. That is, without the presence of a control group, it would be difficult to determine that the observed improved accuracy in writing stemmed solely from WCF—not from other sources. To improve the methodological rigor, it is important to include a control group in WCF research to accurately assess the effectiveness of WCF (Liu and Brown 2015). Kang and Han found that WCF is effective at improving learners' accuracy in their writing when compared with no provision of WCF. Specifically, they reported a medium to large effect based on an immediate posttest ($g = .68$). However, they did not identify clear-cut differences in effect sizes between different types of feedback (e.g., indirect vs. direct and focused vs. unfocused). As Kang and Han noted, one possible reason for the result is that several factors, such as learners' L2 proficiency, might be associated with the efficacy of different types of WCF. For instance, beginning-level learners are not likely to correct their errors based on indirect WCF due to their lack of linguistic knowledge. Relatedly, another notable finding from Kang and Han's analysis is that larger effect sizes were found as the proficiency level went up. The effect of corrective feedback was smaller when it was given to beginners. Additionally, the genre of the writing task was found to moderate the effectiveness of WCF. Specifically, when corrections were given for journal writing, the effect size was significantly lower. This finding might be linked to the private nature of journal writing, as it is normally not supposed to be read and responded to by others. Finally, an instructional setting was another significant moderator. Learners in a second-language setting benefitted from WCF more than learners in a foreign-language setting. As Ferris (2010) pointed out, this difference might be related to the fact that learners in second-language classes tend to spend more time practicing revising and editing based on WCF, compared to those in foreign language classes, where the instructional focus is often on grammar and reading skills. However, the results from Kang and Han's meta-analysis should be interpreted with caution because these analyses were based only on immediate posttests because of the small number of the primary studies that administered delayed posttests. Specifically, only eleven out of 22 studies provided delayed posttest data.

A similar concern regarding a lack of studies reporting long-term effects was pointed out in Liu and Brown's (2015) study, a review of 44 empirical studies on the effectiveness of WCF. They took stock of methodological practices in WCF research and found that only around 30 percent of the studies reviewed adopted delayed posttests. They considered it a severe flaw, since without delayed posttests, it would be extremely difficult to determine whether learning that had taken place was maintained over time. Another drawback Liu and Brown noted in the existing literature concerned the revision requirements. Only half of the studies reviewed were found to require revision based on WCF. The remaining studies only encouraged students to review WCF on their own for less than ten minutes. They claimed that the revision of writing based on corrections is crucial to promote accuracy in writing in the long term. That is, revision requirements are likely to promote learners' accuracy in writing over time. Another methodological flaw plaguing the WCF studies, according to Liu and Brown, was that many studies employed a single-shot design in which feedback was provided on a draft only once. In general, feedback sessions were limited to one or two in almost all of the studies. However, the extended duration of feedback sessions might be a key factor linked to the efficacy of WCF.

To overcome these methodological shortcomings, recent studies on WCF have begun to adopt research designs with a delayed posttest and multiple treatments of feedback over time (e.g., Frear and Chiu 2015, Karim and Nassaji 2018, Rassaei 2019). One of Truscott's (1996) arguments against WCF practices was that feedback is likely to result in pseudo-learning, which is an ephemeral form of knowledge and not linked to consistent accuracy in the long term. The question raised by Truscott—does WCF help students to improve their accuracy over time?—has not yet been scrutinized thoroughly. To answer this question, research is needed to examine the long-term effect sizes of WCF, particularly with regard to whether key variables, such as types of feedback, proficiency levels, the genre of writing, and settings, play significant roles in the accuracy of L2 writing. To contribute to studies on L2 WCF, the present study involved a meta-analytic approach to synthesizing empirical studies examining long-term effects of WCF. The efficacy of WCF has been meta-analyzed (e.g., Kang and Han 2015), but no study to date has directly compared the short- and long-term effects of WCF at the meta-analytic level. This study synthesized WCF studies adopting both immediate and delayed posttests and assessed the short- and long-term effects of WCF. Furthermore, moderator variables were assessed based on both immediate and delayed posttests, and the results were compared to identify factors that affect the short- and long-term effects of WCF. Specifically, the study was guided by the following two research questions:

1. Is there any difference in the short- and long-term effects of WCF on L2 learners' accuracy in writing?
2. What factors modulate the short- and long-term effects of WCF on L2 learners' accuracy in writing?

3. Method

3.1 Literature Search

To identify primary studies, a systematic search was conducted, which consisted of three stages. First, relevant studies were identified through electronic databases, including Linguistics and Language Behavior Abstracts (LLBA), Google Scholar, Educational Resources Information Center (ERIC), PsycInfo, and ProQuest. The following search terms or a combination of them were used: "written corrective feedback," "error correction," "accuracy," "written comments," "response," "revision," "second language," and "foreign language."

Next, the reference sections of each identified study, other previous meta-analyses, and methodological syntheses on this topic (Liu and Brown 2015, Truscott 2007) were examined for potentially relevant studies. Third, to thoroughly canvass recent studies carried out since the previously published meta-analysis (i.e., Kang and Han 2015), we manually scanned abstracts of studies published after Kang and Han's (2015) cutoff date (December 2013) in applied linguistics journals, including *Language Learning*, *TESOL Quarterly*, *Studies in Second Language Acquisition*, *Applied Linguistics*, *Language Teaching Research*, *Foreign Language Annals*, *Journal of Second Language Writing*, and *The Modern Language Journal*.

To decrease the likelihood of publication bias, both published (peer-reviewed) and unpublished theses or dissertations were considered. As for unpublished studies, the search was limited to unpublished Ph.D. dissertations. MA theses were not included in this meta-analysis as in Liu and Brown's (2015) study. The studies identified were evaluated using the inclusion criteria, which are specified in the following section, and the full texts of studies that satisfied the inclusion criteria were retrieved.

3.2 Inclusion/Exclusion Criteria

The timespan of the search was set to range between January 1, 1980 and June 30, 2019. Following Kang and Han's (2015) meta-analysis, the studies were considered for inclusion if they met the following criteria:

1. The study was published in English between 1980 and 2019. Specifically, June 30th of 2019 was the cut-off date on which data collection for the present study was completed. The year 1980 was chosen as the starting year because there are only a few studies on WCF that appeared prior to 1980.
2. The study investigated feedback provided by teachers, not by computers or peers. However, a study on computer-mediated feedback offered by teachers (e.g., Rassaei 2019) was included.
3. The study investigated the efficacy of WCF in improving L2 grammatical accuracy. Therefore, a study that investigated the learning of pragmatic features, such as honorifics (e.g., Cho and Kim 2019), was not included.
4. The study measured students' grammatical accuracy using a writing task rather than editing or revision tasks (e.g., Ashwell 2000, Fathman and Whalley 1990, Ferris and Roberts 2001).
5. The study used a writing task to elicit a sufficiently large amount of learners' written output. For example, in Shintai and Aubrey's (2016) study, participants were required to produce only five sentences and received WCF based on those sentences during the treatment session. Given the limited writing production, the study was excluded.
6. The study examined both short-term and long-term effects of written corrective feedback using both immediate and delayed posttests.
7. The study employed a (quasi-) experimental design that included a comparison or control group.
8. The study's experimental and control groups had at least 10 participants.
9. The study did not have any confounding variables, such as conferences between teachers and students, so that the treatment effects on students' writing performance could be attributable to written corrective feedback.
10. The study reported statistical information necessary to estimate effect sizes (i.e., means, standard deviations, sample sizes, *t*-test)

3.3 Coding

After careful review of each retrieved study, a total of 25 empirical studies were included in this meta-analysis: 12 studies from Kang and Han (2015) and 13 new studies published between 2014 and 2019. The coding scheme was based on existing review studies on WCF (Kang and Han 2015, Liu and Brown 2015), and it is broadly composed of four categories, as shown in Table 1.

Table 1. Coding Scheme

Categories	Variables	Levels
Study Characteristics	Author(s)	Open-ended
	Title	Open-ended
	Publication year	Open-ended
	Type of publication	Peer-reviewed journal, Dissertations, Book chapters
	Setting	Second Language (SL), Foreign Language (FL)
		Open-ended
	First language	Open-ended
	Second language	Open-ended
	Target language	Beginning, Intermediate, Advanced
	L2 proficiency	Children (1-12), Teen (13-17), Adult
	Age	Elementary, Secondary, University, Language program
Feedback	School	
	Type	Direct, Indirect, Metalinguistic explanation, Mixed
	Scope	Unfocused, Focused
	Linguistic focus	Open-ended
	Number of feedback sessions provided	Open-ended
	Genre of writing tasks	Open-ended
Measurement	Revision required	Yes, No
	Outcome measures	Accuracy rate, Error rate
	Interval of treatment and posttest	Open-ended
	Interval of immediate and delayed posttest	Open-ended
Effect size data	Mean and standard deviation	Sample size, Group mean, Standard deviation (for treatment and control groups)

First, studies were coded based on their characteristics, such as author, type of publication, participants' age, L2 proficiency, instructional setting (e.g., Foreign Language or Second Language), etc. In terms of feedback, its (a) type, (b) scope, (c) particular linguistic focus, and (d) number of treatment sessions were coded. As for feedback type, Ellis' (2008) typology of feedback was adopted. As such, feedback was coded as direct if teachers provided corrections on students' errors and as indirect if they only located the presence of students' errors, for example, by underlining or writing a code to indicate its type. For studies with two or more direct feedback conditions (i.e., Bitchener and Knoch 2010, Ekiert and Gennaro 2021), only the condition that was closest to Ellis' definition of direct feedback was included and the data from the other groups were excluded. For instance, Bitchener and Knoch (2010) explored two direct conditions, one which involved direct feedback and another which involved direct feedback + metalinguistic explanation, but only the data from the former condition was coded as direct and included in the calculation for the overall effect of direct feedback. Next, the scope of feedback was classified as either focused or unfocused. Focused feedback targeted only a few grammatical errors, no more than three,

whereas unfocused feedback targeted a wide-range of linguistic structures (Bitchener and Ferris 2012). Finally, the total number of feedback sessions provided during treatment was recorded.

Similar to previous review studies, each study was coded for (a) the genres of writing tasks used and (b) the requirements for students to revise in response to written feedback. As for the latter variable, it was not addressed in the previous meta-analysis, but was included in the present study because this variable could play a crucial role in moderating the efficacy of WCF considering that requiring students to revise can improve the salience of feedback by drawing students' attention to it (Bitchener and Ferris 2012, Guenette 2007).

As for outcome measures, they were categorized as gauging either accuracy rate or error rate. Measures of accuracy rate focused on L2 learners' improvement in the correct use of grammatical forms or structures in writing, whereas measures of error rate focused on learners' reduction of errors.

In addition, in order to investigate the durability of the effects of WCF, the interval between treatment and immediate posttest and the interval between immediate and delayed posttest were coded. When more than one delayed posttest was administered in a study, the first delayed posttest was used for the analysis of the delayed effect. Each variable mentioned above was considered as a possible moderator. For effect size calculation, each study was coded based on sample size, mean, and standard deviation for both the control and treatment groups.

3.4 Data Analysis

All statistical analyses were conducted via the Comprehensive Meta-Analysis (CMA, version 2) software developed by Borenstein, Hedges, Higgins and Rothstein (2005). Following recent L2 meta-analyses (e.g., Kang and Han 2015, Kang et al. 2019), Hedges's g was adopted for effect size estimation because unlike Cohen's d , it amends bias stemming from small sample sizes ($n < 20$) (Lipsey and Wilson 2001). As for effect size magnitude, Oswald and Plonsky's (2010) SLA-specific benchmarks for between-group effects were adopted to interpret effect sizes. Following their reference criteria, 0.4 was regarded as small, 0.7 medium, and 1.00 large.

In addition, a random-effects model was chosen rather than a fixed-effects model to estimate the overall effect sizes. The fixed-effects model was not adopted because it assumes that the effect sizes in all studies included are equal. This assumption is unlikely to be true in educational studies given various student populations and educational settings. Due to the methodological diversity of the identified samples, a random-effects model was found more suitable for the current meta-analysis (Borenstein, et al. 2011).

Statistical analyses involved the following two steps: effect size calculation and moderator analysis. Effect sizes were calculated by comparing treatment groups with comparison or control groups. To retain statistical independence, more than one effect size from a single study was not included in the analysis (Lipsey and Wilson 2001). Instead, effect sizes from each study were averaged to yield overall effects. For instance, when a study examined two types of feedback, two effect sizes could be calculated, but they were aggregated to generate only a single effect size for the study.

As for multiple effect sizes generated from immediate and delayed posttests in a single study, individual effect sizes were calculated separately for each outcome measure administered at immediate and delayed posttest. That is, the effect sizes for immediate and delayed posttests were calculated independently from a single study and were treated as separate entities so that the overall effect size for immediate posttests could be compared to the average effect size of delayed posttests.

As a final step, moderator analysis was run by calculating the Q -statistic to estimate effect size heterogeneity. The goal of the moderator analysis was to find out whether certain moderator variables led to variability among the effect sizes of the included studies. Specifically, Q between (Qb) tests were adopted to pinpoint moderator

variables. The results of each significant Qb value for each variable indicate whether or not the variable is a significant moderator. When moderator analyses were conducted, methodological issues also arose regarding multiple effect sizes from a single study. In order to maintain statistical independence, only one effect size was included from a single study when each moderator was investigated. Examples include studies that assessed various types of feedback. Sheen et al. (2009) investigated direct + focused and direct + unfocused feedback. Since both types of feedback were direct, a composite effect size was computed by averaging the effect sizes, and the composite effect size was included in the moderator analysis of types of feedback (direct vs. indirect). In addition, due to the reasons mentioned above, random-effect size estimates were selected to conduct moderator analyses. However, as suggested by Borenstein et al. (2011), a fixed-effects model was used if there were less than five studies ($k < 5$) involved in the analysis of a certain moderator variable.

4. Results

4.1 Descriptive Analysis

In total, 25 studies satisfied the inclusion criteria. 12 studies, which were included in our previous meta-analysis, were published between 1980 and 2013, and 13 studies were published between 2014 and 2019. These figures suggest that there has been a considerable increase in the number of studies investigating the long-term effects of WCF during the last ten years, as presented in Figure 1. Of 25 studies, 21 (84 %) were published in journals, and 4 studies (16 %) were unpublished doctoral dissertation studies. As for the setting, 12 of the studies (48 %) were conducted in a second language (SL) context, and 13 (52 %) were carried out in a foreign language (FL) context.

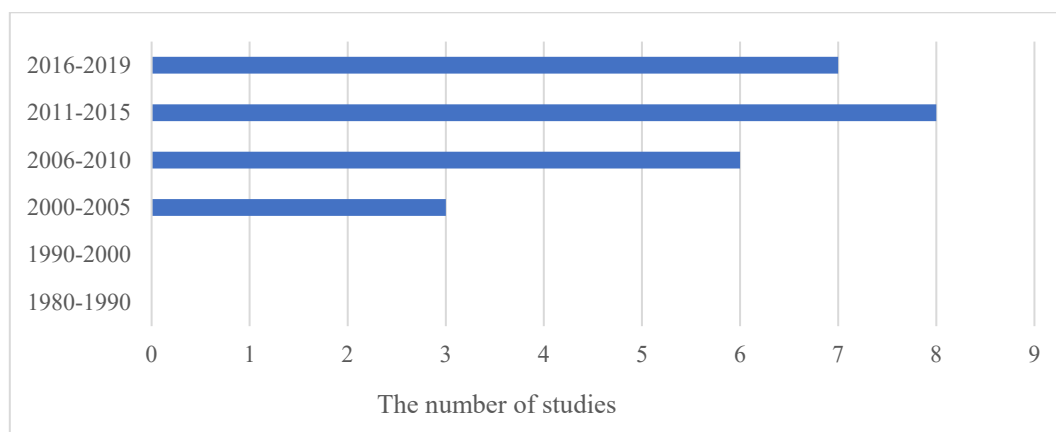


Figure 1. The Number of Studies Investigating Both Short- and Long-term Effects of WCF

Table 2 summarizes more information on study characteristics. As shown in Table 2, a variety of participants' L1s were reported, including, Chinese, Greek, Japanese, Vietnamese, Korean, Persian, Arabic, and Spanish. However, about half of the studies were conducted with a group of participants whose L1 is different ($k = 12$ or 48 %). With regard to the target language involved, the majority of the studies ($k = 22$ or 88 %) investigated English, and only three studied another language (German, Spanish, and French). 15 studies involved intermediate learners and two were conducted with advanced learners. Also, 23 studies (92 %) were carried out with adults, and only two studies examined teenagers. Among the 25 included studies, most of the studies investigated participants

enrolled in either a university ($k = 12$ or 48 %) or a language program ($k = 10$ or 40 %), and only three studies (12 %) studied secondary school students.

Table 2. Study Characteristics

Variables	Levels	%	k
Learners' L1	Chinese	8	2
	English	4	1
	Greek	4	1
	Japanese	8	2
	Vietnamese	4	1
	Korean	4	1
	Mixed	48	12
	Arabic	12	3
	Persian	4	1
	Spanish	4	1
L2 (Target language)	English	88	22
	German	4	1
	Spanish	4	1
	French	4	1
L2 proficiency	Advanced	8.0	2
	Intermediate	60	15
	Not specified	32	8
Age	Adults	92	23
	Teens	8	2
Instructional status	Language program	40	10
	Secondary	12	3
	University	48	12
Linguistic focus of feedback	English articles	48	12
	Comprehensive (no focus)	24	6
	Tense	12	3
	Pronoun	8	2
	Case	4	1
	Mixed (less than four)	4	1
Number of feedback sessions	1	24	6
	2~3	64	16
	>3	12	3
Time interval between immediate and delayed posttest	1~2 weeks	36	9
	3~4 weeks	28	7
	>4 weeks	36	9

As for the linguistic focus of feedback, roughly half of the studies ($k = 12$) targeted articles. Close to 20 % ($k = 6$) provided more comprehensive feedback (e.g., unfocused) that focused on a variety of grammatical errors. The remaining studies targeted tense, pronouns, or a limited number of grammatical errors (less than four).

With regard to the number of feedback sessions provided, 76 % of the studies provided more than a single treatment, but 24 % of the studies included only a single provision of feedback. Lastly, the time interval between immediate and delayed posttests ranged from 1 week to 10 months. However, on average, delayed posttests were administered roughly 6 weeks after the immediate posttest. Table 2 presents more descriptive information on the types and scope of feedback and the genres of writing tasks used.

4.2 Quantitative Meta-analysis

To detect the existence of publication bias, several analyses were conducted. Publication bias may occur if the sample of studies retrieved for a meta-analysis mainly includes primary studies with significant results. Simply put, it is caused by the selective inclusion of studies (Borenstein et al. 2011). To check the presence of publication bias, funnel plots were first constructed. As seen in Figure 2, the funnel plots appeared symmetric around the midpoint, representing no existence of publication bias. In addition, a trim-and-fill analysis (Duval 2005) was conducted. The analysis adjusts for publication bias by re-imputing an overall effect size based on the number of possibly neglected studies because of their insignificant results. The adjusted effect size, shown as an open diamond at the bottom, was the same with the observed effect size, represented by a closed diamond. These results suggest that publication bias does not exist in the present meta-analysis.

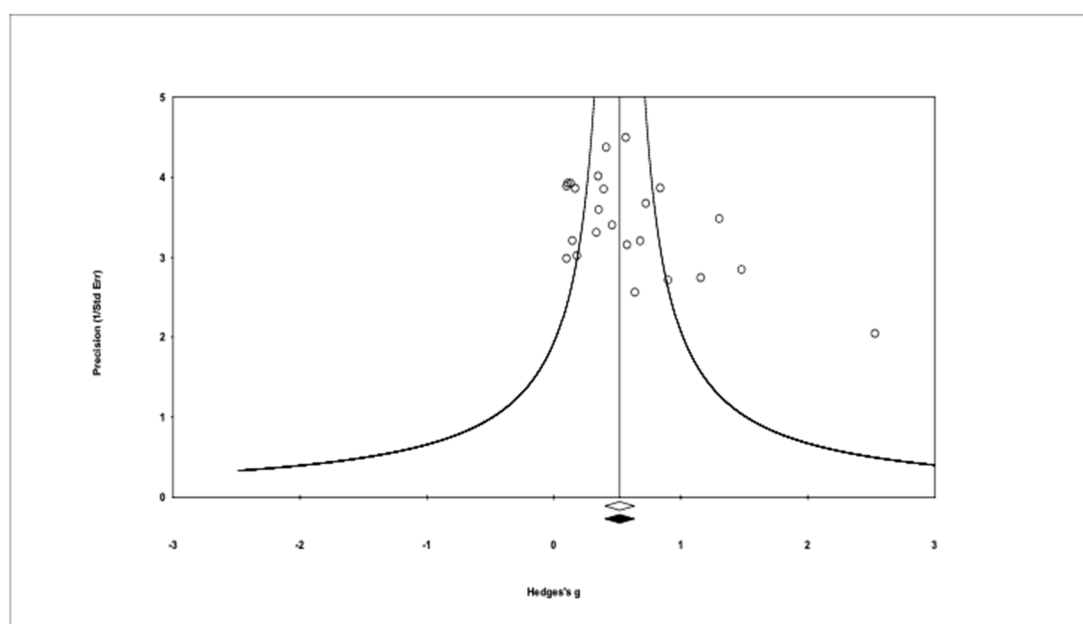
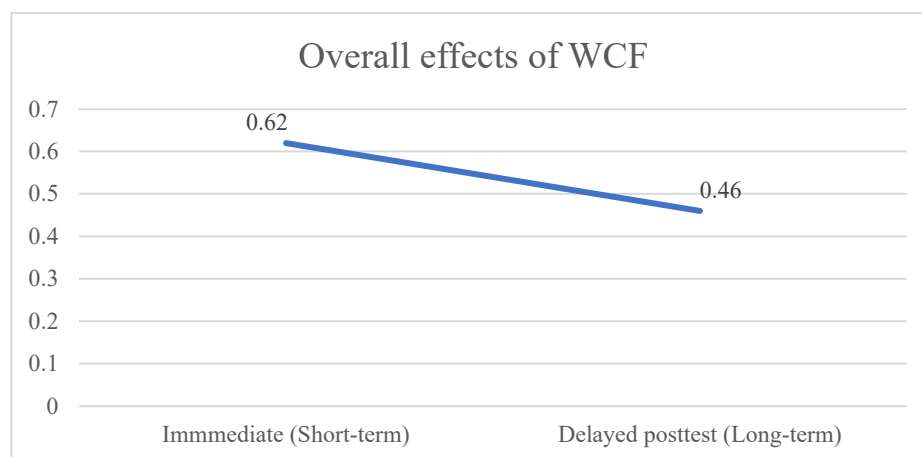


Figure 2. Funnel Plot of Standard Error by Hedges' g

After publication bias was explored, the overall effects of WCF were computed. In general, WCF was found to result in positive effects on L2 learners' written accuracy. Specifically, as shown in Table 3, WCF led to medium effects on learner accuracy on immediate posttests ($g = 0.62$). However, a small effect size was found on delayed posttests ($g = 0.46$), as shown in Figure 3. That is, the effects of WCF fell slightly from immediate to delayed posttests.

Table 3. Effects Sizes of Written Corrective Feedback at Posttest and Delayed Posttest

	<i>n</i>	<i>k</i>	Mean <i>ES</i> (<i>g</i>)	<i>SE</i>	95 % CI	
					Lower	Upper
Overall						
Immediate posttest	25	25	0.62	0.105	0.414	0.827
Delayed posttest	25	25	0.46	0.098	0.270	0.654

**Figure 3. Overall Effects of WCF**

4.3 Moderator Analysis

Table 4 presents the results of the moderator analyses. As for most of the moderator variables, immediate effects were larger than delayed effects. First, age did not moderate the short-term and long-term effects of WCF. The variation between adults and teens was not statistically significant both at immediate, $Q(1) = 0.05, p > .05$, and delayed posttests, $Q(1) = 0.06, p > .05$. In terms of the setting, the effectiveness of WCF was not different between FL and SL contexts. There was only a minor difference between them at immediate posttests ($Q(1) = 0.01, p > .05$), but the difference was more noticeable at delayed posttests. The effects of WCF provided in SL contexts ($g = 0.37$) were found to be greater than that given in FL contexts ($g = 0.59$) even though the difference was not statistically significant, $Q(1) = 1.26, p > .05$.

With regard to proficiency, a larger effect size was found in studies involving advanced learners than those targeting intermediate learners. No statistical difference, however, was identified between the two levels of learners, both at immediate, $Q(1) = 2.32, p > .05$, and delayed posttests, $Q(1) = 0.80, p > .05$. As for educational status, studies conducted in secondary schools yielded a smaller effect size compared to those carried out in language programs or universities, but the difference among language programs, universities, and secondary schools was somewhat negligible, and the variable was not statistically significant both at immediate, $Q(2) = 2.01, p > .05$, and delayed posttests, $Q(2) = 0.17, p > .05$.

Table 4. Summary of Moderator Analyses

Moderator	Immediate Posttest					Delayed Posttest				
	<i>k</i>	<i>g</i>	Lower 95 % CI	Upper 95 % CI	<i>Qb</i>	<i>k</i>	<i>g</i>	Lower 95 % CI	Upper 95 % CI	<i>Qb</i>
Age					0.05					0.06
Adults	23	0.51	0.39	0.63		23	0.44	0.32	0.55	
Teens	2	0.58	-0.04	1.20		2	0.36	-0.26	0.97	
Setting					0.01					1.26
FL	13	0.56	0.30	0.82		13	0.37	0.09	0.64	
SL	12	0.54	0.28	0.80		12	0.59	0.30	0.89	
Proficiency					2.32					
Advanced	2	0.83	0.36	1.23		2	0.67	0.18	1.16	0.80
Intermediate	15	0.45	0.30	0.56		15	0.44	0.29	0.58	
School					2.01					
Language program	10	0.60	0.41	0.79		10	0.41	0.23	0.60	0.17
Secondary	3	0.30	-0.09	0.69		3	0.38	-0.002	0.78	
University	12	0.50	0.34	0.66		12	0.46	0.30	0.62	
Type of feedback					8.62**					5.27*
Direct	16	0.44	0.30	0.59		16	0.41	0.27	0.56	
Indirect	3	1.03	0.67	1.39		3	0.88	0.51	1.25	
Scope of feedback					0.81					0.73
Focused	6	0.55	0.29	0.80		6	0.41	0.18	0.64	
Unfocused	6	0.39	1.47	0.63		6	0.63	0.20	1.05	
Revision required					0.10					0.002
No	9	0.57	0.32	0.81		19	0.47	0.26	0.68	
Yes	6	0.51	0.29	0.73		6	0.48	-0.06	0.99	
Genre of writing task					9.56*					15.5**
Academic writing task	6	0.38	0.16	0.59		6	0.48	0.26	0.70	
Narrative	5	0.43	0.18	0.68		5	0.20	-0.05	0.45	
Picture description	2	0.72	0.52	0.91		12	0.65	0.47	0.83	
Text reconstruction	2	0.24	-0.12	0.58		2	-	-0.38	0.32	
							0.03			
Number of feedback provided					0.07					3.51
1	6	0.52	0.30	0.75		6	0.48	0.25	0.71	
2~3	6	0.51	0.37	0.65		16	0.48	0.34	0.63	
>3	3	0.56	0.20	0.92		3	0.13	-0.22	0.48	

As to the type of feedback, indirect feedback was found to be more effective than direct feedback. Significant differences among the different feedback types were observed in the immediate, $Q(1) = 8.62, p < .01$, and delayed overall effects, $Q(1) = 5.27, p < .05$. Regarding revision, there was no significant difference in effect sizes between studies that asked for revision and those that did not require revision on immediate, $Q(1) = 0.10, p > 0.05$, and delayed posttests, $Q(1) = 0.002, p > .01$. The non-significant Qb statistic results suggested that this variable did not modulate the short-term and long-term effects of WCF.

Next, the genre of writing tasks was investigated as a moderator variable. A similar pattern of effects was found between immediate and delayed tests. The effect size of WCF was largest when picture description tasks were used, followed by academic writing tasks, then narratives, and finally, text reconstruction tasks. This variation among the different genres of writing tasks was statistically significant at immediate, $Q(3) = 9.56, p < .05$ and delayed posttests, $Q(3) = 15.5, p < .01$.

Finally, the difference in terms of the number of feedback sessions provided was not found on immediate, $Q(2) = 0.07, p > .05$, and delayed measures, $Q(2) = 3.51, p > .05$. Regardless of the number of feedback sessions, small to medium effect sizes were found on immediate and delayed posttests.

4. Discussion

This study reports the results of a meta-analysis of WCF studies adopting both immediate and delayed posttests. Previous studies have mainly documented the short-term effects of WCF. However, thanks to a growing body of research employing both short- and long-term effects, this study could directly compare short- and long-term effects of WCF and investigate how a host of moderator variables function differently depending on time points (e.g., immediate posttest vs. delayed posttest).

In terms of the difference in the short- and long-term effects of WCF on L2 learners' accuracy in writing, WCF was found to lead to medium effects on learner accuracy on immediate posttests ($g = 0.62$). The moderate effect sizes on immediate posttests were also identified in prior work (Kang and Han 2015), providing additional evidence for the effectiveness of WCF. However, in this study, a smaller effect size was found on delayed posttests ($g = 0.46$). Although there is no existing meta-analysis reporting the long-term effects of WCF, the finding is not at odds with meta-analytic findings in other areas of instructed SLA, which shows that the effects of instruction on L2 learning taper off over time (Goo et al. 2015).

The second research question addressed in this meta-analysis pertains to factors moderating the overall effects of WCF. Among the nine moderator variables investigated, types of feedback and genre of writing tasks were significant. As for types of feedback, indirect feedback ($g = 1.39$) was found to have a larger effect on L2 accuracy than direct feedback ($g = 0.59$), as revealed by immediate outcome measures. Indirect feedback ($g = 1.25$) also appeared to have a longer-lasting effect on accuracy, as evidenced by delayed posttests, than direct feedback ($g = 1.25$). The finding is inconsistent with that of Kang and Han's (2015) meta-analysis, which failed to detect differences between the two types of feedback. As a possible reason for the result, Kang and Han pointed out that the type of feedback variable might work in concert with other variables, such as proficiency. For example, beginning-level learners can benefit more from direct correction than from indirect feedback due to their lack of grammar knowledge and ability to self-correct their errors. Kang and Han's explanation is applicable to the present meta-analysis because the participants of the primary studies in the data set were mainly intermediate learners, who had sufficient L2 knowledge to utilize indirect WCF. This characteristic of the population might have resulted in higher effect sizes for indirect feedback. However, the current finding that indirect feedback seems to be superior to direct feedback in the short and long term for L2 learning confirms the prediction in the literature regarding their relative effectiveness. L2 researchers have maintained that indirect feedback may facilitate progress toward developing accuracy because it encourages learners to self-correct their errors and to engage in a deeper level of language processing. Thus, indirect feedback would be helpful for L2 learners to internalize correct forms in the long run (Kang and Han 2021). Direct corrections, on the other hand, deprive L2 learners of such learning opportunities by instantly providing correct forms. Yet, it must be noted that caution should be exercised in

interpreting the current finding, given the small number of studies on indirect WCF ($k=3$) included in the analysis. Future empirical studies are necessary to elucidate the relative effectiveness of direct and indirect feedback and its interaction with time.

The genre of writing task was also found to moderate both the short- and long-term effects of WCF. While Kang and Han (2015) compared the overall effect sizes of studies based only on three genres of writing tasks (compositions, journals, and letters), in this meta-analysis, comparisons were made among four categories—academic writing tasks, narratives, picture descriptions, and text reconstruction—due to a different subset of primary studies. In this meta-analysis, the largest effect sizes were found in picture description tasks for immediate and delayed posttests. The finding might be related to the unique attribute of the picture description task. In general, picture prompts require learners to give a written account of certain objects, scenes, activities portrayed in them. The unambiguous contexts might make WCF more salient to the learners. On the other hand, when WCF was offered in text reconstruction, the effect sizes were considerably low. Its overall effect size on the immediate test ($g = 0.24$) fell significantly further on the delayed posttest ($g = -0.03$). The finding might be associated with the contrived nature of the text reconstruction task. The task compels learners to read a given L2 text closely so that they can reconstruct it later. Students might not be motivated enough to review WCF in their reconstructed text because they might perceive errors as the result of the memory test. However, to date, there have been no empirical studies directly exploring the relationship between the genre of writing tasks and the effects of WCF. As Kang and Han suggested, L2 teachers should understand that the genre of writing tasks can play a role in the effects of WCF and that certain types of writing tasks, such as text reconstruction, might be less responsive to feedback.

Besides the genre of writing tasks and types of feedback, however, other variables were not found to contribute to the effect size variance. A somewhat unexpected finding from the moderator analyses is that while revisions are recommended in the literature in response to WCF to promote its positive effects, no differences in effect sizes were found based on revision requirements. Requesting revision is likely to lead to increased attention to corrections by requiring learners to process them rather than just looking through them (Liu and Brown 2015). Even though revision is considered crucial to facilitate grammatical accuracy in writing, it is unclear whether it can promote the effectiveness of WCF. For example, recently, Ekanayaka and Ellis (2020) tested the utility of revision by comparing students who had a chance to revise with those who did not. Their findings confirm the pedagogical assumption, but interestingly, the WCF alone was also found effective even when it was not accompanied with a requirement to revise. Ekanayaka and Ellis's study is one of the few empirical studies on the issue, and therefore, more empirical studies are necessary to elucidate the role of revision in the effectiveness of WCF.

Another unanticipated finding is that there were no clear differences based on the number of feedback sessions provided. That is, more feedback did not lead to higher gains in writing accuracy. It seems puzzling that longer-term WCF treatments did not result in larger effect sizes. Nonetheless, the number of feedback sessions was not a significant moderator variable in previous analyses, either. Kang and Han (2015) noted that short-term intervention tends to have a narrower focus, and is therefore more noticeable. Another possibility for this counterintuitive finding might be related to an artifact of the coding category adopted in the study. Many existing studies have employed a single-shot design, where feedback was offered only on a single draft, and in almost all of the existing studies, only up to three or four feedback sessions were provided. Due to the limited number of feedback sessions provided in the primary studies, for the current analysis, the variable was coded into three categories: (i) once, (ii) twice or three times, and (iii) more than three times. It is possible that the effects of the number of treatment are not appreciable when there are only small differences prompted by one or two WCF sessions. The finding calls for more studies administering more sessions of WCF treatments to determine long-term effects of WCF. The

benefits of extended WCF periods are well demonstrated in theoretical terms. For example, the U-shaped learning pattern in SLA indicates that language development is nonlinear and that grammatical accuracy is developed over time (Gass and Selinker 2008). That is, initial exposure to WCF might enable learners to make corrections and use forms accurately. Nonetheless, they might regress or temporarily fail to use the correct forms before they eventually internalize them. Hence, L2 learning takes a long time, and, thus, it is only by exploring WCF over time that we may have a better and fuller understanding of its effects on L2 learning. If the goal is to measure the long-term development in L2 learning as a result of WCF, research designs based on a single WCF session might be misleading. Therefore, more studies that use feedback in multiple rather than a couple of sessions are needed.

5. Conclusion and Future Directions

Whereas the effects of WCF on L2 accuracy in writing have been documented at the meta-analytic level, the results are limited to short-term effects. The present meta-analysis was focused WCF studies employing immediate and delayed posttests to evaluate and compare the short- and long-term effects of WCF directly. The study found that WCF yielded a positive effect on L2 accuracy, compared to the no-WCF condition. The overall effect of WCF was moderate ($g = 0.62$) on immediate posttests but fell within the small-to-moderate range ($g = 0.46$) on delayed posttests. As for moderator variables, they functioned in a similar fashion depending on the two test time points. Specifically, the genre of writing tasks and types of feedback were the significant mitigating factors on the immediate and delayed posttests.

This meta-analysis contributes to ongoing WCF research by investigating the short- and long-term effects of WCF on L2 accuracy in writing, but the sample of the studies was restricted to 25. Although 13 more studies employing delayed posttests are included in the current dataset compared to Kang and Han's (2015) meta-analysis, it is still limited in terms of a firm determination of whether the effects of WCF are maintainable. Therefore, more research into the long-term effects of WCF is needed to systemically test its role in L2 learners' accuracy in writing. In addition to studies gauging the lasting effects of WCF via delayed posttests, another methodological improvement required for experimental studies is to expand the number of sessions of WCF provided. Unless the number of WCF sessions provided is extended greatly, it would be difficult to decide whether Truscott's (1996) argument against WCF still holds. Additionally, future studies may help expand our understanding of the long-term effects of WCF by increasing the interval between the immediate and delayed posttests. In general, the delayed posttests in WCF studies were only modestly delayed—administered roughly two weeks after the immediate posttests. To adequately rebut Truscott's argument that knowledge acquired from WCF can “disappear in a matter of months” (p. 346), delayed posttests should be expanded substantively in future research. This could help future meta-analyses aimed toward accruing more accurate knowledge about the long-term effects of WCF.

References

**References marked with an asterisk indicate studies included in meta-analysis.*

Ashwell, T. 2000. Patterns of teacher response to student writing in a multi-draft composition classroom: Is content feedback followed by form feedback the best method? *Journal of Second Language Writing* 9, 227–257.

*Bitchener, J. 2008. Evidence in support of written corrective feedback. *Journal of Second Language Writing* 17, 102–118.

- Bitchener, J. and D. Ferris. 2012. *Written Corrective Feedback in Second Language Acquisition and Writing*. New York: Routledge.
- *Bitchener, J. and U. Knoch. 2010. Raising the linguistic accuracy level of advanced L2 writers with written corrective feedback. *Journal of Second Language Writing* 19, 207–217.
- *Bitchener, J. and U. Knoch. 2010a. The contribution of written corrective feedback to language development: A ten-month investigation. *Applied Linguistics* 31, 193–214.
- *Bitchener, J., S. Young and D. Cameron. 2005. The effect of different types of corrective feedback on ESL student writing. *Journal of Second Language Writing* 9, 227–258.
- Borenstein, M., L. V. Hedges, J. P. T Higgins and H. R. Rothstein. 2011. *Introduction to Meta-analysis*. West Sussex, UK: Wiley.
- Borenstein, M., L. V. Hedges, J. P. T Higgins and H. R. Rothstein. 2005. *Comprehensive Meta-analysis* (version 2). Englewood, NJ: Biostat.
- Cho, H. and Y. Kim. 2019. Learning Korean honorifics through individual and collaborative writing tasks and written corrective feedback. *Applied Linguistics Review* 13(1), 19–47.
- Duval, S. 2005. The trim and fill method. In H. R. Rothstein, A. J. Sutton and M. Borenstein eds., *Publication Bias in Meta-analysis: Prevention, Assessment and Adjustments*, 127–144. Chichester, UK: Wiley.
- Ekanayaka, W. I. and R. Ellis. 2020. Does asking learners to revise add to the effect of written corrective feedback on L2 acquisition? *System* 94, 1–12.
- *Ekiert, M. and K. di Gennaro. 2021. Focused written corrective feedback and linguistic target mastery: Conceptual replication of Bitchener and Knoch (2010). *Language Teaching* 54, 71–89.
- Ellis, R. 2008. A typology of written corrective feedback types. *ELT Journal* 63(2), 97–107.
- *Ellis, R., Y. Sheen, M. Murakami and H. Takashima. 2008. The effects of focused and unfocused written corrective feedback in an English as a foreign language context. *System* 36, 353–371.
- Fathman, A. K. and E. Whalley. 1990. Teacher response to student writing: focus on form versus content. In Kroll, B. ed., *Second Language Writing: Research Insights for the Classroom*, 178–190. Cambridge: Cambridge University Press.
- *Fazio, L. L. 2001. The effect of corrections and commentaries on the journal writing accuracy of minority- and majority-language students. *Journal of Second Language Writing* 10(4), 235–249.
- Ferris, D. and J. Hedgcock. 2014. *Teaching L2 Composition*. New York: Routledge.
- Ferris, D. 2006. Does error feedback help student writers? New evidence on the short- and long- term effects of written error correction. In Hyland, K. and F. Hyland, eds., *Feedback in Second Language Writing: Contexts and Issues*, 81–104. Cambridge: Cambridge University Press.
- Ferris, D. 2010. Second language writing research and written corrective feedback in SLA. *Studies in Second Language Acquisition*, 32, 181–201.
- Ferris, D. and B. Roberts. 2001. Error feedback in L2 writing classes: How explicit does it need to be? *Journal of Second Language Writing* 10, 161–184.
- *Frear, D. and Y. Chiu. 2015. The effect of focused and unfocused indirect written corrective feedback on EFL learners' accuracy in new pieces of writing. *System* 53, 24–34.
- Gass, S. M. and L. Selinker. 2008. *Second Language Acquisition: An Introductory Course*. New York: Routledge.
- Goo, J., G. Granena, Y. Yilmaz and M. Novella. (2015). Implicit and explicit instruction in L2 learning: Norris & Ortega (2000) revisited and updated. In P. Rebuschat, ed., *Implicit and Explicit Learning of Languages*, 443–482. Amsterdam: John Benjamins.

- Guénette, D. 2007. Is feedback pedagogically correct?: Research design issues in studies of feedback on writing. *Journal of Second Language Writing* 16(1), 40–53.
- Kang, E. Y. and Z-H Han. 2015. The efficacy of written corrective feedback in improving L2 written accuracy: A meta-analysis. *The Modern Language Journal* 99, 1–18.
- Kang, E. Y. and Z-H Han. 2021. Written Corrective Feedback: Short-Term and Long-Term Effects on Language Learning. In R. M. Manchón and C. Polio, eds., *The Routledge Handbook of Second Language Acquisition and Writing*, 213–225. New York: Routledge.
- Kang, E., S. Sok, and Z. Han. 2019. Thirty-five years of ISLA on form-focused instruction: A meta-analysis. *Language Teaching Research* 23(4), 428–453.
- *Karim, K., and H. Nassaji 2018. The revision and transfer effects of direct and indirect comprehensive corrective feedback on ESL students' writing. *Language Teaching Research* 24(4), 519–539.
- Karim, K., and H. Nassaji. 2019. The effects of written corrective feedback. *Instructed Second Language Acquisition* 3(1), 28–52.
- Lipsey, M. and D. Wilson. 2001. *Practical Meta-analysis*. Thousand Oaks, CA: SAGE.
- Liu, Q. and D. Brown. 2015. Methodological synthesis of research on the effectiveness of corrective feedback in L2 writing. *Journal of Second Language Writing*, 30, 66–68.
- *López, M. B., E. V. Steendam, D. Speelman and K. Buyse. 2018. The Differential Effects of Comprehensive Feedback Forms in the Second Language Writing Class. *Language Learning* 68(3), 813–850.
- *Mawlawi Diab, N. 2015. Effectiveness of written corrective feedback: Does type of error and type of correction matter? *Assessing Writing* 24, 16–34.
- *Mawlawi Diab, N. 2016. A comparison of peer, teacher and self-feedback on the reduction of language errors in student essays. *System* 57, 55–65.
- *Mubarak, M. 2013. *Corrective Feedback in L2 Writing: A Study of Practices and Effectiveness in the Bahrain Context*. Unpublished doctoral dissertation, University of Sheffield, United Kingdom.
- Oswald, F. L., and L. Plonsky. 2010. Meta-analysis in second language research: Choices and challenges. *Annual Review of Applied Linguistics* 30, 85–110.
- *Rassaei, E. 2019. Computer-mediated text-based and audio-based corrective feedback, perceptual style and L2 development. *System* 82, 97–110.
- *Rummel, S. and J. Bitchener. 2015. The effectiveness of written corrective feedback and the impact Lao learners' beliefs have on uptake. *Australian Review of Applied Linguistics* 38(1), 66–84.
- *Sheen, Y. 2007. The effect of focused written corrective feedback and language aptitude on ESL learners' acquisition of articles. *TESOL Quarterly* 41, 255–283.
- * Sheen, Y. 2010. Differential effects of oral and written corrective feedback in the ESL classroom. *Studies in Second Language Acquisition* 32(2), 203–234.
- *Sheen, Y., D. Wright and A. Moldawa. 2009. Differential effects of focused and unfocused written correction on the accurate use of grammatical forms by adult ESL learners. *System* 37, 556–569.
- *Shintani, N. and R. Ellis. 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.
- *Shintani, N., R. Ellis and W. Suzuki. 2014. Effects of written feedback and revision on learners' accuracy in using two English grammatical structures. *Language Learning* 64(1), 103–131.

- *Stefanou, C. and A. Révész. 2015. Direct written corrective feedback, learner differences, and the acquisition of second language article use for generic and specific plural reference. *The Modern Language Journal* 99(2), 263–282.
- *Suh, B. R. 2010. *Written Feedback in Second Language Acquisition: Exploring the Roles of Type of Feedback, Linguistic Targets, Awareness, and Concurrent Verbalization*. Unpublished doctoral dissertation, Georgetown University, U.S.A.
- Truscott, J. 1996. The case against grammar correction in L2 writing classes. *Language Learning* 46, 327–369.
- Truscott, J. 2007. The effect of error correction on learners' ability to write accurately. *Journal of Second Language Writing* 16, 255–272.
- *Wagner, J. 2016. *An Explanation of the Effectiveness of Written Corrective Feedback in Second-Language Acquisition*. Unpublished doctoral dissertation, George Mason University, U.S.A
- *Zhang, X. 2017. Reading–writing integrated tasks, comprehensive corrective feedback, and EFL writing development. *Language Teaching Research* 21(2), 217–240.

Examples in: English
Applicable Languages: English
Applicable Level: Tertiary