The Modern Language Journal

The Efficacy of Written Corrective Feedback in Improving L2 Written Accuracy: A Meta-Analysis

EUNYOUNG KANG
Columbia University
Teachers College
525 W 120th Street
New York, NY
Email: kang2@tc.columbia.edu

ZHAOHONG HAN
Columbia University
Teachers College
525 W 120th Street
New York, NY
Email: han@tc.columbia.edu

Written corrective feedback has been subject to increasing attention in recent years, in part because of the conceptual controversy surrounding it and in part because of its ubiquitous practice. This study takes a meta-analytic approach to synthesizing extant empirical research, including 21 primary studies. Guiding the analysis are two questions: Does written corrective feedback help to improve the grammatical accuracy of second language writing? What factors might mitigate its efficacy? Results show that written corrective feedback can lead to greater grammatical accuracy in second language writing, yet its efficacy is mediated by a host of variables, including learners' proficiency, the setting, and the genre of the writing task.

Keywords: written corrective feedback; meta-analysis; second language writing

WRITTEN CORRECTIVE FEEDBACK IS A COMmon instructional strategy, considered essential and nonnegotiable by many, to help second language (L2) learners improve their writing effectiveness (Ferris, 2010). Teachers, in general, spend a great deal of time providing various kinds of corrections (e.g., grammar, spelling) to varying extents (e.g., correcting every error or selectively a few), in the conviction that such feedback is necessary to improving students' written accuracy and their ability to write. However, written corrective feedback is still a divisive issue among researchers (Ferris, 2010, 2012). Conceptual differences aside, empirical studies have led to inconclusive, if not conflicting, findings. In particular, research results have pointed to two contentious issues: (a) when and how written correct feedback works and (b) what type of feedback strategy is effective.

The Modern Language Journal, 99, 1, (2015)
DOI: 10.1111/modl.12189
0026-7902/15/1–18 $1.50/0

Still, the fact that a considerable number of studies have been conducted attests, on the one hand, to an ongoing interest in written corrective feedback and suggests, on the other hand, that it is now possible to apply a meta-analytic approach to help resolve some lingering issues. Such efforts are, in fact, deemed worthwhile and desirable, as they may elucidate a seemingly confusing domain of study, identify gaps and conflicts in extant findings, and, last but not least, provide for some practical understanding of written corrective feedback that may guide teachers' practice. A meta-analysis, by design, aims at deriving so-called aggregate findings by combining and integrating results from multiple primary studies.

In this article, we report on a meta-analysis of extant written corrective feedback studies, with a view to both deriving a broad understanding of the efficacy of written corrective feedback and identifying and assessing the impact of factors that may mitigate its efficacy.

BACKGROUND

Written corrective feedback, which involves attempts to rectify errors—primarily grammatical

errors—in L2 learners' writing, is a common pedagogical strategy in L2 writing classrooms. Despite its popularity, views differ on whether written corrective feedback is helpful to L2 learners' grammatical development. Truscott (1996, 1999, 2007) infamously claims that error correction[1] may help learners revise a subsequent draft but does little to improve their grammar as they attempt a new piece of writing. In other words, Truscott recognizes the potential value of written corrective feedback in making learners conscious of their errors and leading them to correct surface deviances from the putative target forms, but he challenges its role in fundamentally altering their knowledge representation. Truscott (1996) goes so far as to claim that "grammar correction has no place in writing courses and should be abandoned" (p. 328). Contrary to this view,[2] Ferris (1999, 2003) maintains that written corrective feedback, if unambiguous and consistent, can lead to fundamental improvement in learners' grammatical accuracy. In their attempt to reconcile the controversy, Hyland and Hyland (2006) have conducted an extensive narrative review, noting, inter alia, that "it is difficult to draw any clear conclusions and generalizations from the literature as a result of varied populations, treatments and research designs" (p. 84) and, importantly, concluding that we are still falling short of a clear understanding of the effects of written feedback on the development of L2 learners' written accuracy.

Consequently, the topic of written corrective feedback has continued to garner research interest and attention, vis-à-vis two central concerns: developing a generic understanding of the efficacy of written corrective feedback and identifying potential mediating factors. In pursuing an understanding of the latter, for instance, several studies have examined the efficacy of written corrective feedback as a function of its scope—focused versus unfocused—in improving L2 learners' overall accuracy (e.g., Ellis, 2009; Ellis et al., 2008; Sheen, Wright, & Moldawa, 2009; for oral feedback, see Doughty & Varela, 1998; Han, 2002). Ellis (2009) reports that focused feedback that targets fewer grammatical errors is more effective than unfocused feedback that targets more errors, arguing that the former focuses, rather than disperses, learner attention, thereby enabling learners to more readily recognize the difference between their own use and the target-like use of forms. Similarly, a study by Farrokhi and Sattarpour (2012) yields superior effects for focused feedback over unfocused feedback.

Another alleged mediating factor is the type of feedback—direct versus indirect feedback. Whereas indirect feedback only signals the locus of an error, direct feedback explicitly corrects it by not only signaling its locus, but also providing its correct counterpart. Ellis et al. (2008) report that direct feedback (focused or unfocused) has a durable positive impact on the grammatical accuracy of EFL learners' writing. Similarly, Chandler (2003) observes that direct feedback outstrips indirect feedback. Lalande (1982), on the other hand, finds that indirect feedback engages students more than does direct feedback, and it promotes longer term development. Evidently, the jury is still out on the type of feedback as a substantive variable. Many believe that the efficacy of either direct or indirect feedback may, actually, be contingent on its interaction with another factor, learner proficiency (see, e.g., Ellis, 2009; Ferris & Roberts, 2001). Ellis (2009), for example, asserts that direct feedback is more useful for beginners, who still need explicit guidance to extend their linguistic repertoire, while indirect feedback is sufficient for advanced learners, who, due to their high proficiency, have the capacity to discern errors for themselves.

In addition, researchers have also explored the quantity and timing of corrective feedback as well as the genre of writing as potential mitigating forces. These efforts notwithstanding, what remains unknown yet critical to understanding written corrective feedback is how these potential factors operate, and when they do, what magnitude of influence they carry.

Meta-analyses would be an effective way to address these questions, as we have seen in the fruitful application of several meta-analyses to feedback research to date (see, e.g., Biber, Nekrasova, & Horn, 2011; Li, 2010; Mackey & Goo, 2007; Russell & Spada, 2006; Truscott, 2007). A quantitative treatment of a delimited body of research, a meta-analysis can effectively distinguish different types of feedback, assess their respective magnitudes of impact, and isolate factors and determine their contributions to the impact variance.

Previous meta-analyses have primarily addressed the impact of oral corrective feedback (see Table 1). In their meta-analysis of 27 published studies, Mackey and Goo (2007) found a moderate to large effect size ($d = 0.71$ based on immediate posttests and $d = 1.09$ based on short term delayed posttests) for interactional feedback such as clarification requests and recasting. The Russell and Spada (2006) meta-analysis was a more general

TABLE 1
Previous Meta-Analyses on L2 Corrective Feedback

Authors	Cutoff Year	Source of Feedback	Mode of Corrective Feedback	N Studies (*k*)	Outcomes	Effect Size (ES) (Cohen's *d*)	ES Overall
Russell & Spada (2006)	2004	Teacher Native speaker (NS) Peers Researcher Other NS	Oral + Written	15	L2 grammar learning	0.91 (oral) 1.31 (written)	1.16
Truscott (2007)	2006	Teacher	Written	13	Accuracy in L2 writing	0.148 (uncontrolled studies) −0.155 (controlled studies)	
Li (2010)	2007	Teacher Native speaker Computer	Oral	33	L2 grammar acquisition		FE: 0.61/ RE: 0.64
Biber et al. (2011)	2007	Teacher Peers Computer	Written + Oral	23	L1 + L2 writing development	Pre/posttest design 0.86 (oral) 0.68 (written) Treatment/control design 0.84 (oral) 0.40 (written)	

Note. FE: Fixed-effects model, RE: Random-effects model.

study of corrective feedback, including 15 studies on oral or written feedback. As shown in Table 1, an overall effect size of 1.16 was reported for both oral and written feedback, and an independent effect size of 1.31 specifically for written corrective feedback, indicating a large effect of corrective feedback in general, and of written feedback specifically, on L2 grammar learning.

In contrast, Li (2010), including mostly oral corrective feedback studies, found less substantial effects, $d = 0.61$ and $d = 0.64$ respectively, on fixed and random effects models. Li's analysis was notably more sophisticated than that of its predecessors, in that it included both published and unpublished studies (22 and 11), thereby addressing the publication bias variable (Norris & Ortega, 2000) while investigating an array of moderator variables that putatively affect the effectiveness of corrective feedback. Two findings stand out from Li's analysis, among others. First, implicit feedback has a longer term effect than explicit feedback. Second, corrective feedback is more effective (i.e., has a larger effect size) in a foreign language setting than in a second language setting. However, because the meta-analysis focused largely on oral corrective feedback, it remains an open question whether or not the findings extrapolate to written corrective feedback.

A recent meta-analysis by Biber and colleagues (2011) examined the effects of various types of corrective feedback on the quality of students' writing, with 23 studies published between 1982 and 2007 serving as its database. The analysis delved into the source (teacher, peer, tutor, student, or computer), the focus (e.g., grammar, vocabulary, spelling, organization, content, mechanics, rhetorical organization), and the type (e.g., direct comment, editing code, rating) of feedback. A moderate to large effect of written corrective feedback was found: $d = 0.4$ for studies with a treatment/control design and $d = 0.68$ for studies with a pre- and posttest design. However, these results must be taken with a grain of salt, because the database included both first language (L1) and L2 studies. Thus, the efficacy of L2 written corrective feedback has yet to be ascertained.

To date, Truscott (2007) has been the only published meta-analysis that is devoted entirely to written corrective feedback research. The study reports a small effect size of 0.148 for six uncontrolled studies and −0.155 for six controlled studies, concluding that written corrective feedback plays only a tangential role in improving L2 written accuracy. Pioneering and informative as it is, this meta-analysis is notably narrow in scope in that it examined only 12 published

empirical studies. Importantly, since Truscott (2007), the general database for written corrective feedback research has substantially expanded, with more than a dozen additional studies appearing in the L2 literature. The time, therefore, seems ripe both to update the findings of Truscott (2007) and to gain new insights about the efficacy of L2 written corrective feedback through, for example, examining mitigating variables, something Truscott did not do in his meta-analysis.

The goal of the present meta-analysis was to build on and complement the existing research on written corrective feedback. Three questions were addressed:

RQ1. Is written corrective feedback generically effective for improving L2 written accuracy?
RQ2. If so, which type of written corrective feedback is more effective?
RQ3. What factors mitigate the efficacy of written corrective feedback?

METHOD

Identifying Primary Studies

Our first step was to build an adequate corpus of primary studies through the following procedures:

1. Consulting three online databases—the Educational Information Resource Center (ERIC), LLBA (Linguistics and Language Behavior Abstract), and ProQuest—using these search terms or combinations thereof: *error correction*, *written corrective feedback*, *response*, *comment*, *writing*, *composition*, *writing instruction*, *writing process*, *writing skills*, *editing*, *revision*, *writing improvement*, *foreign language*, and *second language*.
2. Performing manual searches of the following journals that provide online access: *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*.
3. Using Google Scholar to search for additional studies or verify studies.
4. Consulting the reference sections of relevant book chapters, published meta-analyses, and narrative reviews (e.g., Bitchener, 2012; Hyland & Hyland, 2006) as additional sources of potential studies. Once the prospective studies were identified, the article abstracts were carefully screened, based on a set of inclusion/exclusion criteria, resulting in the final sample of 22 primary studies.

Inclusion/Exclusion Criteria

Studies had to pass the following screening criteria for inclusion in the present meta-analysis:

1. The study had to examine L2 written corrective feedback provided by a teacher rather than by peers or a computer.
2. The study had to use the grammatical accuracy of students' original writing as a dependent variable, rather than that demonstrated on editing or revision tasks (e.g., Ashwell, 2000; Fathman & Whalley, 1990; Ferris & Roberts, 2001).[3]
3. The study had to be published in 1980 or later, pursuant to a practice adopted in previous meta-analyses. For the current study, the cutoff date was set at December 2013.
4. The study had to use an experimental or quasi-experimental design that included a control group or a comparison group. In particular, the study had to contrast at least two groups: a treatment group (with feedback) and a control or comparison group (without feedback).[4] This criterion was based on a broad agreement among L2 written feedback researchers that a control group is needed in written feedback research (for discussion, see Ferris, 1999, 2004; Truscott, 2004).
5. The study had to be reported in English.

When the same experimental result was reported in multiple published studies, as in Bitchener and Knoch (2009, 2010a), only one was included in the meta-analysis. In addition, the present analysis excluded those studies (a) that did not provide sufficient statistics for calculation of effect sizes and/or (b) whose goal was to improve oral production through written corrective feedback. Studies were also excluded if the effects of feedback could not be isolated from that of other treatments such as conferences (e.g., Polio, Fleck, & Leder, 1998; Sheppard, 1992).[5]

Data Coding

The final sample of 22 studies (18 published studies and 4 unpublished doctoral dissertations[6]) was coded following the suggestions of Lipsey and Wilson (2001). Specifically, four categories of variables were coded, as shown in

TABLE 2
Coding Scheme

Main Category	No. of Variables	Variables
Study Characteristics	8	Study ID Author Publication year Type of publication Mean age of participants Setting L2 proficiency Instructional status
Research Design	4	Type of research Interval between pretest and treatment Interval between treatment and posttest Interval between immediate and delayed posttest
Treatment	7	Number of target features Scope of feedback Type of feedback Number of feedback sessions Genre of writing task Target language Outcome measure
Effect Size	8	Effect size type Total sample size Treatment group size Treatment group mean Treatment standard deviation Control group size Control group mean Control group standard deviation

Table 2 (see also Appendix A for detail), and these included Study Characteristics, Research Design, Treatment, and Effect Size.

The first category, Study Characteristics, subsumes eight variables, including the study ID, the author, the year of publication, the type of publication, the mean age of participants, the study setting, participants' L2 proficiency, and the instructional status.

The second category, Research Design, includes the type of research, the interval between pretest and treatment, the interval between treatment and posttest, and the interval between immediate and delayed posttest.

The third category, Treatment, contains seven variables: the number of target features, the scope of feedback, the type of feedback, the total number of feedback sessions, the genre of the writing task, the target language, and the outcome measure.

The fourth category, Effect Size, includes the type of effect size, the total sample size, the size of the treatment group, the mean of the treatment group, the standard deviation of the treatment group, the size of the control group, the mean of the control group, and the standard deviation of the control group.

Moderator Variables

Drawing on insights from previous studies, especially Russell and Spada (2006), in the present study nine variables[7] were gleaned from 19 primary studies as moderator variables, among them four from the first coding category (Study Characteristics)—year of publication, setting, instructional status, and proficiency level—and five from the third category (Treatment)—scope, type of feedback, number of feedback sessions, genre of writing task, and target language (see Appendix B).

For the variables associated with Study Characteristics, the setting variable was coded into foreign language (i.e., where learners have little to zero exposure to the target language outside the classroom) or second language (i.e., where learners have exposure to the target language both in and out of the classroom). The instructional status was coded into elementary school, secondary school, language institute, or university. Learners' proficiency was coded into

beginning, intermediate, or advanced, as reported in the original studies.

For the set of Treatment variables, first, the scope variable was coded in terms of focused versus unfocused feedback. Focused feedback involves selective correction of only a limited range of, and typically predetermined, error types (Ellis et al., 2008), while unfocused feedback refers to correcting every grammatical error in the learner's writing. Second, the type of feedback is coded into direct or indirect feedback. Direct feedback involves providing the correct counterparts of incorrect forms, while indirect feedback offers typographic cues such as underlining and circling the incorrect forms but has learners figure out their correct counterparts for themselves (Ferris, 2006). Third, the number of feedback sessions was coded into 1, 2, 3, or more than 3 sessions. Fourth, the genre of the writing tasks used during Treatment was coded into composition, journal, letter, or narrative. Finally, the target language of written corrective feedback was coded into English or "other" (all other languages). We deemed the selected moderator variables worthwhile to examine for several reasons. First, as suggested in previous meta-analyses (Li, 2010; Russell & Spada, 2006), the effectiveness of written corrective feedback may vary depending on the settings and levels of learners' proficiency. Second, research has indicated that the scope and type of feedback may have a different effect on individual learners (Brown, 1994; Ellis, 2009; Ferris, 2002). Third, investigating the effect sizes of different types of written feedback may provide useful information for teachers in determining which type of feedback to deploy. Lastly, the effectiveness of written corrective feedback may vary depending on the number of feedback sessions, the genres of the writing tasks, and the language medium of feedback.

Reliability of Coding

To ensure coding reliability, all of the primary studies were coded multiple times until saturation was reached. Then, an outside coder was brought in to code a random subset of five studies. The overall interrater agreement rate was 97%, and the differences were resolved through discussion.

Analysis

In this meta-analysis, effect sizes were calculated based on outcome measures of written grammatical accuracy. All measures were reported as continuous variables, generally in the form of error and accuracy rate of students' writing. With regard to effect size statistics, Hedges's *g*, a conservative version of Cohen's *d*, was adopted because unlike Cohen's *d*, which has a small sample bias, Hedges's *g* corrects for biases due to small sample sizes (Lipsey & Wilson, 2001). Effect sizes were calculated using the means, standard deviations, and sample sizes of the treatment and control groups in each study. When studies did not report standard deviations, *t* values were used to calculate an estimate of Hedges's *g*. All effect sizes were calibrated using a professional meta-analysis program called Comprehensive Meta-Analysis (CMA) (Borenstein et al., 2005).

For our purposes, the random effects model was adopted, which is premised on the assumption that the true effect size varies across studies. This assumption is deemed appropriate, given that the magnitude of the impact of a given treatment can be different from study to study, due to variations in the age of participants, the setting, and so on (Borenstein et al., 2011).

For the effect size calculation, only a single effect size was used per construct for each study rather than several effect sizes to avoid violation of the assumptions of statistical analysis (Lipsey & Wilson, 2001). Due to their inclusion of multiple treatments, subgroups, and outcome measures, many studies contributed more than one effect size. To deal with this issue, multiple effect sizes associated with a single construct within a study were averaged.[8] For instance, Ellis et al. (2008) investigated the effects of two different types of corrective feedback: direct/focused vs. direct/unfocused. In this case, because the two types of feedback were both instantiations of a superordinate construct—corrective feedback—the effect sizes were averaged in the final analysis for the overall effect. Similarly, Bitchener, Young, and Cameron (2005) employed multiple outcome measures such as the accuracy rates of preposition, tense, and article usage, all of which are tied to a larger construct of grammatical accuracy. In this case, effect sizes from the multiple measures were averaged. Moreover, given that not every study conducted a delayed posttest, only the immediate posttest measures were calculated for the overall effect.

Once the overall effect size of written corrective feedback was estimated, the heterogeneity across effect sizes was checked using *Q* statistics, which determines whether "variability across effect sizes is greater than expected from sampling error" (Lipsey & Wilson, 2001, p. 133). A significant *Q*

statistic would mean that the effect sizes are not homogeneous. Heterogeneity is likely to arise from variations in the type of written corrective feedback, outcome measure, and participants across the studies. Rejecting the null hypothesis of homogeneity warrants the test for moderator effects, which serves to identify potential variables that may mitigate the effects of written corrective feedback. Although the Q statistic and its p-value provide information on the statistical significance of heterogeneity, they do not show the exact amount of dispersion. Thus, in order to explain the degree of heterogeneity, T^2 and I^2 statistics were additionally computed. T^2 is the amount of true heterogeneity, and I^2 reflects "the proportion of observed dispersion that is real" (Borenstein et al., 2011, p. 125).

To examine the sources of heterogeneity across the effect sizes, an analysis was subsequently performed of moderator variables. In this part of the analysis, only the studies that reported any of the moderator variables noted earlier in the Moderator Variables section were included. For example, if a study did not provide information on the learners' proficiency level, the effect size data from that particular study were excluded from the analysis on the proficiency variable, though still included in the overall mean effect size calculation and in other moderator analyses. Furthermore, the *one study, one effect size* principle was upheld for the moderator analyses. For example, in Van Beuningen et al. (2008), the impact of direct and indirect feedback was examined by including two treatment groups (direct and indirect) and a single comparison group (self-correction). In a case like that, the effect size of indirect feedback could be calculated as the difference between the direct feedback and comparison group, while that of direct feedback could be computed as the difference between the same comparison and the indirect feedback group. The problem there, though, is that the same control group would be contributing to more than one effect size, and this would violate the assumption of data independence (Borenstein et al., 2011). As a result, Van Beuningen, de Jong, and Kuiken's (2008) study was purged from the moderator analysis of type of feedback.

Potential moderator variables were evaluated by conducting Q Between (Qb) tests to determine the extent to which these variables account for the variance in the effects of written corrective feedback. Qb values, therefore, serve as indicators of whether effect sizes significantly vary across the moderator variables.

RESULTS

Overall Effect of Written Corrective Feedback

Based on a sample of 22 studies, written corrective feedback appears to have a moderate to large effect on the grammatical accuracy of L2 students' writing (Hedges's $g = .68$, $SE = .13$, $CI = .42 \sim .93$, $p < .0001$). There was also significant heterogeneity across the effect sizes of the studies ($Q = 87.180$, $p < .0001$). The variance of true effects T^2 was .28, and I^2 was 74.77%, suggesting that 74.77% of the observed heterogeneity might be due to real differences between studies, not a random error. These results justified the subsequent analysis of moderator variables to account for variance in effect sizes across the studies. A summary of the effect sizes of the individual studies is provided in Appendix C.

While, as noted earlier, the delayed posttests were excluded from the overall effect analysis so as to preserve the independence of effect sizes, there were 11 studies that did report a delayed posttest and were analyzed for their delayed effect sizes. Some of these studies reported more than one delayed posttest, in which case the first delayed posttest was chosen for analysis. In general, the delayed posttests were administered at least 15 days after the corrective feedback treatment (see Appendix B). Their mean effect size was as follows: $g = .682$, $SE = .095$, $CI = 0.496 \sim 0.869$, $p < .0001$, indicating moderate effects for learning retention. Caution is, nonetheless, required when interpreting these results due to the small sample size.

Outliers and Publication Bias

To discover an outlier, Z scores, which are the standardized estimates of effect sizes, were inspected. As it turned out, only one outlier was found, the Van Beuningen, de Jong, and Kuiken (2012) study. The Z scores showed that the effect size of this study was five standard deviations above the overall effect size. When the outlier was excluded, the overall effect size slid to $g = .54$ ($CI = .35 \sim .72$) from $g = 0.68$ ($CI = .35 \sim .74$). The study was consequently excluded from all analyses. To determine whether there was a publication bias (i.e., whether effect sizes happened to come equally from studies with large or small sample sizes) in the sample, a funnel plot of effect sizes across 21 studies was checked. The estimates of the effect sizes of studies with the largest samples are shown at the top of the funnel, and those with the small sample sizes at the

bottom (Duval & Tweedie, 2000). If the funnel plot roughly appeared symmetric, it would be indicative of the unlikelihood of a publication bias, meaning that the meta-analysis had captured all the relevant studies (Borenstein et al., 2011). Conversely, an asymmetrical funnel shape would indicate the existence of missing studies in the dataset. For instance, if it showed a lack of studies in the bottom left corner, that would suggest that those studies were not included in the dataset because they presented no significance or did not exhibit the positive effect expected (Duval & Tweedie, 2000). As shown in Figure 1, the studies included in the present meta-analysis are dispersed equally on either side of the overall effect, thereby indicating that the sample is free from publication bias.

Because the funnel plot is a subjective visual tool for detecting publication bias, a trim-and-fill (Duval, 2005) analysis was also carried out to locate missing studies, if any, that may have been neglected due to publication bias (i.e., studies not included in the meta-analysis due to their non-significant results), to add them to the analysis, and to re-impute an overall effect size (Borenstein et al., 2011). It was found that no value of the overall effect size was missing. Under the random effects model, the overall effect size for the combined studies was $g = .54$ ($CI = .35 \sim .72$). The trim-and-fill analysis showed that the re-imputed estimate was the same. The results were, therefore, consistent. Furthermore, a fail-safe N was calculated, with the result suggesting that 304 studies would have to be neglected to invalidate a significant effect size result (rejecting the null hypothesis that the effect size is the same as 0.00). This number far exceeded the criterion number (i.e., $5k + 10 = 115$ where $k = 21$ studies; Rosenthal, 1991). The fail-safe N result therefore further confirmed that publication bias was not an issue in the present meta-analysis.

Moderator Analyses

Moderator analyses were performed to find out whether certain variables were associated with differences in the effects of written corrective feedback. For the most part, the random effects models were employed, but when the number of studies in a subgroup turned out to be less than five, the fixed effects models were adopted, as recommended by Borenstein et al. (2011). The analyses were conducted using effect sizes associated solely with the immediate posttests.

The first moderator variable analyzed was the study setting. As shown in Table 3, written corrective feedback was significantly more effective when provided in a second language setting ($Qb[1] = 7.65$, $p = .006$) than in a foreign language setting.

Proficiency was also examined as a moderator variable. Significant differences were found among three levels, the advanced, the intermediate, and the beginning ($Qb[2] = 6.89$, $p = 0.032$).

FIGURE 1
Funnel Plot of Standard Error by Hedges's g (Observed and Imputed)

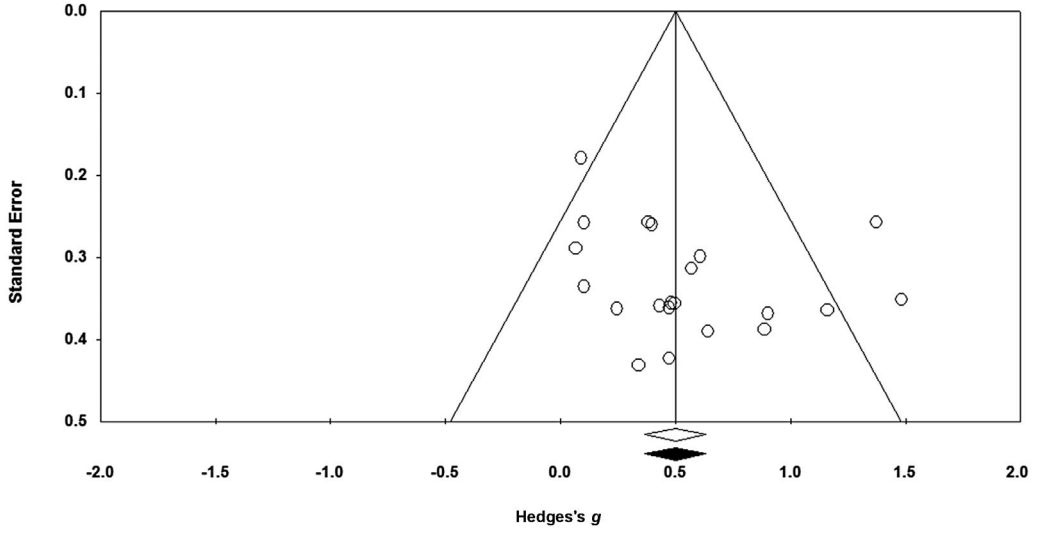

TABLE 3
Summary of Moderator Variables

Moderator Variable	*k*	Effect Size (*g*)	Lower 95% CI	Upper 95% CI	*Qb* Value
Setting					7.65*
Foreign language	6	0.217	−0.003	0.437	
Second language	15	0.664	0.436	0.891	
Proficiency					
Advanced	3	0.726	0.345	1.108	7.04*
Intermediate	12	0.556	0.384	0.729	
Beginning	1	0.089	−0.261	0.440	
School					0.755
Primary	1	0.481	−0.215	1.177	
Secondary	1	0.888	0.127	1.648	
Language institute	7	0.524	0.298	0.749	
University	12	0.468	0.294	0.638	
Scope of feedback					3.366
Focused	10	0.692	0.353	1.030	
Unfocused	9	0.329	0.140	0.518	
Type of feedback					1.58
Direct	12	0.598	0.423	0.774	
Indirect	5	0.361	0.035	0.686	
Treatment sessions					4.67
One shot	5	0.832	0.323	1.341	
2 ∼ 3	6	0.605	0.226	0.983	
> 3	10	0.313	0.125	0.501	
Target language					0.13
English	16	0.583	0.356	0.811	
Non-English	5	0.324	0.074	0.574	
Publication year					5.82*
After 2000	19	0.583	0.435	0.732	
Before 2000	2	0.185	−0.103	0.472	
Genres of writing task					6.2*
Composition	15	0.623	0.459	0.787	
Journal	3	0.228	−0.038	0.494	
Letter	2	0.439	−0.058	0.936	

Note. *k* = number of studies, CI = confidence interval, *$p < .05$.

The effect sizes presented in Table 3 show that beginners did not seem to benefit from written corrective feedback ($g = .09$, SE = .18, $CI = -.26 \sim .44$), but both advanced ($g = .74$, $SE = .2$, $CI = .35 \sim 1.1$) and intermediate learners ($g = .56$, SE = −.09, $CI = .384 \sim .73$) did considerably, with the advanced learners benefitting more than the intermediate learners.

With regard to instructional status as a potential moderator variable, studies conducted in secondary schools exhibited larger mean effect sizes compared to those carried out in universities or language institutes, but the differences were not significant ($Qb[3] = 1.19$, $p = .755$).

Similarly, no significant differences were found between the type of feedback (direct vs. indirect; $Qb[1] = 1.58$, $p = .208$) and the scope of feedback (unfocused vs. focused; $Qb[1] = 3.367$, $p = .067$). Nevertheless, as shown in Table 3, the effect sizes of focused feedback ($g = .69$) and direct feedback ($g = .598$) were higher than their respective counterparts, unfocused feedback ($g = .329$) and indirect feedback ($g = .361$).

Furthermore, an analysis with the number of treatment sessions as a moderator variable showed that studies that provided single sessions of treatment (i.e., written corrective feedback) outperformed those providing multiple sessions. The number of feedback sessions across the studies ranged from one to more than four. As shown in Table 3, the single ($g = .832$, $SE = .26$, $CI = .323 \sim 1.341$) and two or three sessions ($g = .605$, $SE = .193$, $CI = .226 \sim .983$) resulted in higher effect sizes than the more than three sessions ($g = .313$, $SE = .1$, $CI = .125 \sim .501$). However, these differences were not statistically significant ($Qb[2] = 4.667$, $p = .097$).

Also analyzed as a potential predictor of effect size variation was the target language of the study. It was found that, although the effect size for studies concerning L2 English ($g = .583$, $SE = .116$, $CI = .356 \sim .811$) was larger than that concerning other languages ($g = .324$, $SE = 0.128$, $CI = .074 \sim .574$), the difference was not statistically significant ($Qb[1] = 2.259$, $p = .133$).

The timing of publication of the studies was also analyzed as a potential moderator variable. As Table 3 illustrates, studies published after the year 2000 showed a significantly higher effect size ($g = .587$, $SE = .076$, $CI = .435 \sim .732$) than that of the studies published before 2000 ($g = .185$, $SE = .147$, $CI = -.103 \sim .472$), $Qb[1] = 5.818$, $p = .016$).

Finally, the effect of different genres of the writing tasks was examined as a possible moderator variable. Learners showed significant improvements in their written accuracy following feedback provided to them on compositions ($g = .623$, $SE = .084$, $CI = 0.459 \sim 0.787$). Letter writing ($g = .44$, $SE = .254$, $CI = -.058 \sim .936$), on the other hand, yielded a moderate effect size, while journal writing exhibited a small effect size ($g = .228$, $SE = .136$, $CI = -.038 \sim .494$). The differences in the magnitude of effects among the different genres were significant ($Qb[2] = 6.22$, $p = .045$).

DISCUSSION AND CONCLUSION

The first question addressed in this meta-analysis was whether written corrective feedback is effective in the development of L2 learners' written accuracy. The present meta-analysis yielded an overall effect size of $g = .54$, which, according to Cohen (1988), is moderate and which indicates that written corrective feedback does have a substantive effect on L2 written accuracy. According to Cohen (1988), the effect size is low if g varies below 0.2, medium if around 0.5, and large if more than 0.8. However, Cohen's (1988) rule of thumb is often criticized because it does not take into account the context of the specific research domain (Lipsey & Wilson, 2001). Oswald and Plonsky (2010) propose that, within the context of SLA research, 0.40 should be interpreted as small, 0.70 as medium, and 1.0 as large. Based on this SLA-specific benchmark, the effect size found in the present meta-analysis is considered small to moderate. Still, this result seems strongly at odds with that reported in Truscott (2007), which estimated an overall effect size of −0.155 for controlled studies and 0.148 for uncontrolled studies. Given that this meta-analysis included only controlled studies, the overall effect size of $g = .54$ is starkly distinct from that of Truscott (2007) ($d = -.155$). The discrepancy might be due to differences in the selection and inclusion criteria adopted respectively in the present study and in Truscott (2007). For one thing, Truscott selected only published studies based on narrative reviews (Ferris, 1999, 2003, 2004; Truscott, 1996, 1999), whereas the present study included both published and unpublished studies. Moreover, the present study incorporated studies that were conducted after Truscott's cut-off year, 2006,[9] which may have contributed to the potency of the average effect size. While differing from Truscott (2007), the results of the present meta-analysis align well with the Biber et al. (2011) meta-analysis, thus providing additional support for the utility of written corrective feedback.[10]

The second question addressed in the present meta-analysis concerns the efficacy of different types of written corrective feedback, in particular, the indirect vs. the direct. The analysis did not find a clear-cut difference between the two. One possible reason could be that rather than operating alone to temper the effects of written corrective feedback, the type of feedback variable might do so in tandem with other factors such as learners' proficiency in the target language. A number of researchers, including James (1998), have asserted that the indirect approach is more likely to have long-term positive effects on students' accuracy since it requires learners to self-discover the correct forms (Li, 2010). However, Ferris (2002) has maintained that there are several cases in which teachers should provide more direct feedback on errors. She contends that learners benefit more from direct correction when they are at the beginning level of proficiency, when they do not have enough linguistic knowledge to self-correct. Brown (1994) holds a similar view, according to which lower level students, because they are still struggling with understanding discrete forms, are not likely to correct their errors even when the errors are explicitly pointed out to them by teachers. Brown argues that it might be better for these students to have opportunities to practice editing by recopying teacher-provided correct forms. However, to date, there have been no available empirical studies directly investigating the relationship between the proficiency level and the effects of different types of feedback. It is worth mentioning in passing that in her recent narrative review, Ferris (2010) points out an interesting divide: SLA research favoring direct feedback and second

language writing research favoring indirect feedback.

The third question investigated in the present meta-analysis concerns factors influencing the overall effects of written corrective feedback. Among the nine moderator variables analyzed, proficiency stood out as the strongest mitigator of effect size: Larger effect sizes appeared as the proficiency level went up, and the effect of corrective feedback was negative when it was given to beginners. This set of results reinforces the notion that developmental readiness should be considered when providing feedback (Pienemann, 1998). But it must be noted, nevertheless, that the unequal sample size in the present analysis may have skewed the result. The sample in question, as shown in Table 3, included four studies involving advanced learners, twelve involving intermediate learners, but only one (Semke, 1984) involving beginners. Therefore, the finding must perforce be taken with caution.

A somewhat ambivalent finding from our analysis is that learners in a second language setting tend to benefit from written feedback more than learners in a foreign language setting. This finding is at variance with previous studies on oral corrective feedback showing exactly the opposite: that learners in foreign language contexts notice feedback better during communicative interaction, and tend to be more apt at adjusting their output accordingly than learners in second language contexts (e.g., Sheen, 2004). Li's (2010) meta-analysis on oral corrective feedback affirms this finding, reporting higher effect sizes for the foreign language setting than for the second language setting. The disparity might be due to the greater salience of written corrective feedback. Unlike oral feedback, written corrective feedback clearly indicates the presence of errors. Researchers have pointed out that learners often misunderstand oral corrections, such as taking a recast as a signal of agreement rather than correction, rendering the feedback futile (see, e.g., Lyster & Ranta, 1997). Relatedly, due to its greater explicitness than oral feedback, written corrective feedback stands a better chance of being noticed by learners with or without much metalinguistic awareness (Ashwell, 2000). Overall, that written corrective feedback is more potent than oral corrective feedback has been confirmed in Russell and Spada's (2006) meta-analysis.

An alternative explanation for the greater effects of written corrective feedback for the second language setting (e.g., ESL) over the foreign language setting (e.g., EFL) may be, as Ferris (1999, 2010) has noted, that foreign language classes tend to be less concerned about writing as a process and hence that foreign language learners are not keen to revise and edit their writing (Hedgcock & Lefkowitz, 1994, 1996).

Turning now to factors related to the methodological features of written corrective feedback studies, our analysis has revealed that longer term treatment does not yield larger effect sizes than short-term treatment (cf. Li, 2010). In fact, even one-off provisions of written corrective feedback are sufficient to produce large effect sizes. This finding may be accounted for by the fact that short-term treatment tends to have narrower focus and hence is more salient than longer term treatment, which tends to be more extensive in scope and less salient . Also, as pointed out by Li (2010), the duration of treatment may impinge on the "feedback type, intensity of feedback, complexity of linguistic structure, learner differences, and so on" (p. 346). It thus appears that the effects of treatment duration cannot be determined without considering various other design features. Future studies are warranted to delve into the interaction. But before any such attempts are made, it is critical to first identify what the other design features are, an undertaking we assumed in the current meta-analysis.

One of these other features is the genre of the writing task. Our analysis has demonstrated it to be a significant moderator variable, contributing to the effect size variance. In particular, when corrections were provided for journal writing, the effect size was significantly lower. This result might be related to the personal nature of journal writing, as it is usually not intended to be read and responded to by others. Journaling is often used as a task to motivate L2 students to practice writing in a stress-free environment, and the theory of journaling also rejects the idea of error correction because journals are seen as a form of free writing (Walter–Echols, 2008). This is tantamount to suggesting that L2 instructors should be aware that certain types of writing are not as amenable to correction as are other types of writing such as compositions and narratives.

The timing of publication, as our analysis has shown, also contributes to the effect-size variation. Recent studies tend to have significantly larger effect sizes. Of note, studies conducted before 2000 were mostly classroom-based, while the more recent research was in the main experimental or

quasi-experimental. These experimental studies typically provided short-term treatment (correcting once or twice), allowed learners only 10 to 15 minutes to review corrections, and then directed learners to write a new piece of writing to measure the immediate effects of corrective feedback. Clearly, such studies are quick at provoking changes in learners' behavior, yet only longer-term studies can tell whether the changes will endure. This is not to mention a popular weakness of the short-term experimental studies, namely their low ecological validity. It is nothing but a truism that learning a new language requires extensive and sustained meaningful exposure and practice (DeKeyser, 2007; Gass, 2003). Therefore, it only stands to reason that, ultimately, longer term studies should serve as the basis for judging the utility of corrective feedback.

In summary, the present meta-analysis has demonstrated that some variables are connected to the overall efficacy of written corrective feedback. However, as Cooper (1998) has admonished us, the results of the moderator analyses should not be interpreted as implicating casual relationships. Rather, they should be considered as having heuristic values that give confidence in the current analysis, on the one hand, and suggest a need, on the other hand, to subject them to further, systematic study.

LIMITATIONS AND FUTURE DIRECTIONS

Several limitations of this meta-analysis must be noted, as they necessarily mitigate its findings. First of all, this study is based on a restricted sample of 21 studies, not large enough to permit robust comparisons for the moderators considered. As Table 3 shows, the samples for each variable typically have unequal sizes.

A second notable limitation is that the effect sizes are calculated on the basis of immediate posttests. Due to the small number of studies ($k = 11$) employing delayed posttests, the present study was not able to report longer term differences in effect sizes among different moderators. Longitudinal data are as yet lacking but are necessary to determine whether the effects of written feedback are sustainable. Some researchers, among them Guénette (2007), have recommended that future studies be conducted over a longer term—up to several years—to assess how much feedback students need to improve their written accuracy or whether feedback is at all necessary.

Third, constrained by its sample of studies, the present analysis included only a select number of moderator variables; other variables are left out but may equally be worthy of attention, such as learner motivation, attitude toward corrective feedback, and competencies of teachers (Evans et al., 2010), although these have yet to emerge from future primary studies. To date, only a handful of studies have investigated the relative effectiveness of different feedback strategies (Ellis, 2009).[11] As a result, much is still unknown about which types of corrective feedback are more effective. Future research should not only pursue this line of inquiry further but also examine the interaction between different feedback strategies and other variables, including, but not limited to, learners' proficiency and their predilection for explicit and implicit learning (Linck et al., 2013).

It is noteworthy that some of these limitations have come to light also in the meta-analyses on oral feedback (see, e.g., Li, 2010), in particular, the limitations arising from variation in definitions and operationalizations of theoretical constructs and unequal sizes of samples in relation to the variables in question. This shared set of limitations, while underscoring the need for more primary research, speaks to the need to simultaneously explore the effectiveness of oral versus written feedback, importantly, under a uniform set of definitions. Such an effort would help not only to further the general understanding of the role of corrective feedback in L2 learning but also to build a finer grained understanding of the uniqueness of oral versus written feedback. Lalande (1982) has insightfully argued that oral feedback should be selective, but written feedback should be comprehensive. Research comparing focused versus unfocused written corrective feedback appears to have shown that both are effective (see, e.g., Ellis et al., 2008). If this finding can be confirmed in future studies, then it might provide a path to understanding the uniqueness of oral versus written feedback.

The findings from the present meta-analysis, albeit limited and constrained, do have a clear message for L2 writing instructors, that written corrective feedback can improve the grammatical accuracy of student writing.

ACKNOWLEDGMENTS

We are grateful to the anonymous reviewers for their insightful and constructive comments on an earlier draft of this article. Any inadequacies that remain are exclusively our own.

NOTES

[1] A reviewer suggested that "error correction" and "grammar correction" should be differentiated, especially in writing research, because the former can be broader in scope than the latter, including not only grammatical errors but also errors in spelling and mechanics. Such distinction, however, has not been made in the primary studies. In the spirit of being true to the primary research, in this article we use "error correction" and "corrective feedback" interchangeably.

[2] What often evades theoretical and empirical studies citing the written feedback debate is the fact that the conflicting viewpoints espoused by Truscott and Ferris speak fundamentally to an epistemological difference, with Truscott coming from a second language acquisition (SLA) standpoint and Ferris from a second language writing standpoint. The two perspectives are not in unison when it comes to their concerns and questions (for discussion, see Ferris, 2010, 2012).

[3] Students in general tend to pay attention to grammatical accuracy while engaging in revision tasks, so these types of tasks were deemed different in nature from producing a new piece of writing. Truscott (1996, 1999, 2007) argues that corrective feedback may help improve the grammatical accuracy of only a subsequent draft, but not of a new piece of writing. To investigate the validity of this claim, the current meta-analysis included studies that measured students' accuracy in a new piece of writing, and excluded those that used revision tasks, including the dictogloss task (e.g., Shintani, Ellis, & Suzuki, 2014) which involves revision or editing.

[4] A reviewer noted that Chandler's (2003) study, which was included in our sample, did not strictly meet this requirement because both of his groups received feedback. We included it only because Chandler considered it controlled, since students in the comparison group were not required to attend to the feedback they were given until after the study was over.

[5] One reviewer commented that holding conferences with students to talk about the feedback they are given is a normal, appropriate, and maybe even necessary part of the feedback process and thus should not be perceived as some "other treatment" that can get confounded with feedback. Though sympathetic with this view, we, nevertheless, excluded the two studies to ensure greater homogeneity within our dataset, our goal, as it may be recalled, being to investigate the efficacy of L2 written corrective feedback.

[6] The Semke (1980, 1984) study was reported both in a dissertation and a journal article. However, we chose to include the dissertation because it reported more statistical details that we needed for our meta-analysis.

[7] They were identified as factors influencing the overall effect size of written corrective feedback.

[8] A variety of strategies has been recommended for dealing with several effect size estimates from a single study, including choosing the most relevant measure, conducting separate meta-analyses, one for each individual measure, combining all of the effect sizes, using multivariate techniques, and averaging the effect sizes in a single study (Marín–Martínez & Sánchez–Meca, 1999).

[9] Only 5 of the studies in our sample overlap with Truscott's (2007) sample; the remaining 16 are all recent studies.

[10] A reviewer pointed out that the modestly positive results of written corrective feedback reported in the present meta-analysis may have been partly due to the dubious decision to exclude two classroom studies, which were included in Truscott's (2007) meta-analysis, which found correction completely ineffective in one case (Polio, Fleck, & Leder, 1998) and harmful in the other (Sheppard, 1992). However, it may be recalled that the present study also excluded another, much more recent study (Van Beuningen et al., 2012) that reports a very high effect size, in order to eliminate the outliers. The fact that the present analysis included 16 new studies post Truscott (2007) gives credence to our conclusion that recent controlled empirical studies have provided positive evidence for L2 written corrective feedback.

[11] Ellis (2009) proposes a typology of written corrective feedback, which subsumes six different types of strategies: direct correction, indirect correction, focused correction, unfocused correction, electronic correction, and reformulation.

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*, 227–257.

Biber, D., Nekrasova, T., & Horn, B. (2011). *The effectiveness of feedback for L1-English and L2-writing development: A meta-analysis*. ETS Research Report (RR-11-05). Princeton, NJ: ETS.

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

Bitchener, J. (2012). A reflection on 'the language learning potential' of written CF. *Journal of Second Language Writing, 21*, 348–363.

Bitchener, J., Young, S., & Cameron, D. (2005). The effect of different types of feedback on ESL student writing. *Journal of Second Language Writing, 14*, 191–205.

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

Bitchener, J., & Knoch, U. (2009). The value of a focused approach to written corrective feedback. *ELT Journal, 63*, 204–211.

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

Bitchener, J., & Knoch, U. (2010b). Raising the linguistic accuracy level of advanced L2 writers

with written corrective feedback. *Journal of Second Language Writing, 19*, 207–217.

Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. (2005). *Comprehensive meta-analysis* (version 2). Englewood, NJ: Biostat.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2011). *Introduction to meta-analysis*. West Sussex, UK: Wiley.

Brown, A. L. (1994). The advancement of learning. *Educational Researcher*, 4–12.

Chandler, J. (2003). The efficacy of various kinds of error feedback for improvement in the accuracy and fluency of L2 student writing. *Journal of Second Language Writing, 12*, 267–296.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Cooper, H. (1998). *Synthesizing research: A guide for literature reviews*. Thousand Oaks, CA: SAGE.

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

Doughty, C., & Varela, E. (1998). Communicative focus-on-form. In C. Doughty & J. Williams (Eds.), *Focus-on-form in classroom second language acquisition* (pp. 114–138). New York: Cambridge University Press.

Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel-plot–based method of testing and adjusting for publication bias in meta-analysis. *Biometrics, 56*, 455–463.

Duval, S. (2005). The trim and fill method. In H. R. Rothstein, A. J. Sutton, & M. Borenstein (Eds.). *Publication bias in meta-analysis: Prevention, assessment and adjustments* (pp. 127–144). Chichester, UK: Wiley.

Ellis, R. (2009). A typology of written corrective feedback types. *ELT Journal, 63*, 97–107.

Ellis, R., Sheen, Y., Murakami, M., & Takashima, H. (2008). The effects of focused and unfocused written corrective feedback in an English as a foreign language context. *System, 36*, 353–371.

Evans, N. W., Hartshorn, K. J., McCollum, R. M., & Wolfersberger, M. (2010). Contextualizing corrective feedback in second language writing pedagogy. *Language Teaching Research, 14*, 445–463.

Evans, N. W., Hartshorn, K. J., & Strong–Krause, D. (2011). The efficacy of dynamic written corrective feedback for university-matriculated ESL learners. *System, 39*, 229–239.

Farrokhi, F., & Sattarpour, S. (2012). The effects of direct written corrective feedback on improvement of grammatical accuracy of high accuracy of high proficient L2 learners. *World Journal of Education, 2*, 49–56.

Fathman, A. K., & Whalley, E. (1990). Teacher response to student writing: Focus on form versus content. In B. Kroll (Ed.), *Second language writing: Research insights for the classroom* (pp. 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*, 235–249.

Ferris, D. (1999). The case for grammar correction in L2 writing classes: A response to Truscott (1996). *Journal of Second Language Writing, 8*, 1–10.

Ferris, D. (2003). *Response to student writing implications for second language students*. Mahwah, NJ: Lawrence Erlbaum.

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). New York: Cambridge University Press.

Ferris, D. (2010). Second language writing research and written corrective feedback in SLA: Intersections and practical applications. *Studies in Second Language Acquisition, 32*, 191–201.

Ferris, D. (2012). Written corrective feedback in second language acquisition and writing studies. *Language Teaching, 45*, 446–459.

Ferris, D., & Roberts, B. (2001). Error feedback in L2 writing classes: How explicit does it need to be? *Journal of Second Language Writing, 10*, 161–184.

Ferris, D. R. (2002). *Treatment of error in second language student writing*. Ann Arbor: University of Michigan Press.

Ferris, D. R. (2004). The "grammar correction" debate in L2 writing: Where are we, and where do we go from here? (and what do we do in the meantime?). *Journal of Second Language Writing, 13*, 49–62.

Gass, S. (2003). Input and interaction. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 224–255). Malden, MA: Blackwell.

Guénette, D. (2007). Is feedback pedagogically correct? Research design issues in studies of feedback on writing. *Journal of Second Language Writing, 16*, 40–53.

Han, Z. H. (2002). A study of the impact of recasts on tense consistency in L2 output. *TESOL Quarterly, 36*, 543–572.

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*, 84–109.

Hedgcock, J. S., & Lefkowitz, N. (1994). Feedback on feedback: Assessing learner receptivity to teacher response in L2 composition. *Journal of Second Language Writing, 3*, 141–163.

Hedgcock, J. S., & Lefkowitz, N. (1996). Some input on input: Two analyses of student responses to expert feedback on L2 writing. *Modern Language Journal, 80*, 287–308.

Hyland, K., & Hyland, F. (2006). *Feedback in second language writing: Contexts and issues*. Cambridge: Cambridge University Press.

James, C. (1998). *Errors in language learning and use: Exploring error analysis*. London: Longman.

Jhowry, K. (2010). *Does the provision of an intensive and highly focused indirect corrective feedback lead to accuracy?* (Unpublished MA thesis). University of North Texas, Denton, TX.

Kepner, C. G. (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.

Lalande, J. F. (1982). Reducing composition errors: An experiment. *Modern Language Journal, 66*, 140–149.

Li, S. (2010). The effectiveness of corrective feedback in SLA: A meta-analysis. *Language Learning, 60*, 309–365.

Linck, J. A., Hughes, M. M., Campbell, S. G., Silbert, N. H., Tare, M., Jackson, S. R., & Doughty, C. J. (2013). Hi-LAB: A new measure of aptitude for high-level language proficiency. *Language Learning, 63*, 530–566.

Lipsey, M., & Wilson, D. (2001). *Practical meta-analysis.* Thousand Oaks, CA: SAGE.

Lyster, R., & Ranta, L. (1997). Corrective feedback and learner uptake. *Studies in Second Language Acquisition, 19*, 37–66.

Mackey, A., & Goo, J. (2007). Interaction research in SLA: A meta-analysis and research synthesis. In A. Mackey (Ed.), *Conversational interaction in second language acquisition: A series of empirical studies* (pp. 407–453). Oxford: Oxford University Press.

Marín–Martínez, F., & Sánchez–Meca, J. (1999). Averaging dependent effect sizes in meta-analysis: A cautionary note about procedures. *The Spanish Journal of Psychology, 2*, 32–38.

Mubarak, M. (2013). *Corrective feedback in L2 writing: A study of practices and effectiveness in the Bahrain context* (Unpublished doctoral dissertation). The University of Sheffield, Sheffield, UK.

Norris, J. M., & Ortega, L. (2000). Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language Learning, 50*, 417–528.

Oswald, F. L., & Plonsky, L. (2010). Meta-analysis in second language research: Choices and challenges. *Annual Review of Applied Linguistics, 30*, 85–110.

Pienemann, M. (1998). *Language processing and second language development*, Philadelphia/Amsterdam: John Benjamins.

Polio, C., Fleck, C., & Leder, N. (1998). "If I only had more time": ESL learners' changes in linguistic accuracy on essay revisions. *Journal of Second Language Writing, 7*, 43–68.

Rosenthal, R. (1991). *Meta-analytic procedures for social research* (Vol. 6). Thousand Oaks, CA: SAGE.

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). Philadelphia/Amsterdam: John Benjamins.

Semke, H. M. (1980). *The comparative effects of four methods of treating free-writing assignments on the second language skills and attitudes of students in college level first year German* (Unpublished doctoral dissertation). University of Minnesota, Minneapolis, MN.

Semke, H. M. (1984). The effects of the red pen. *Foreign Language Annals, 17*, 195–202.

Sheen, Y. (2004). Corrective feedback and learner uptake in communicative classrooms across instructional settings. *Language Teaching Research, 8*, 263–300.

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*, 201–234.

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*, 556–569.

Sheppard, K. (1992). Two feedback types: Do they make a difference? *RELC Journal, 23*, 103–110.

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*, 286–306.

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*, 103–131.

Sun, S. (2013). *Written corrective feedback: Effects of focused and unfocused grammar correction on the case acquisition in L2 German* (Unpublished doctoral dissertation). University of Kansas, Lawrence, KS.

Truscott, J. (1996). The case against grammar correction in L2 writing classes. *Language Learning, 46*, 327–369.

Truscott, J. (1999). What's wrong with oral grammar correction. *Canadian Modern Language Review, 55*, 437–455.

Truscott, J. (2004). Evidence and conjecture on the effects of correction: A response to Chandler. *Journal of Second Language Writing, 13*, 337–343.

Truscott, J. (2007). The effect of error correction on learners' ability to write accurately. *Journal of Second Language Writing, 16*, 255–272.

Truscott, J., & Hsu, A. Y. (2008). Error correction, revision, and learning. *Journal of Second Language Writing, 17*, 292–305.

Van Beuningen, C. G., de Jong, N. H., & Kuiken, F. (2008). The effect of direct and indirect corrective feedback on L2 learners' written accuracy. *ITL International Journal of Applied Linguistics, 156*, 279–296.

Van Beuningen, C. G., de Jong, N. H., & Kuiken, F. (2012). Evidence on the effectiveness of comprehensive error correction in second language writing. *Language Learning, 62*, 1–41.

Walter–Echols, E. (2008). Journaling as writing practice, reflection, and personal expression. In *CamTESOL Conference on English Language Teaching: Selected Papers, 4*, 120–131.

Wilson, D. B., & Lipsey, M. W. (2001). *Practical meta-analysis.* Thousand Oaks, CA: SAGE.

APPENDIX A

Coding Sheet for the Meta-Analysis Study [Variable Names in Brackets]

Study Characteristics	1. Study ID number [STUDYID] 2. Author 3. Publication year [PUBYear] 4. Type of publication [PUBType] 1. Journal article 2. Thesis or dissertation 3. Book/book chapter 4. Conference proceeding 5. Other (specify): 5. Mean age of sample [Mage] 1. Unspecified 2. Specified 6. Setting [Study Setting] 1. Second language 2. Foreign language 7. L2 proficiency level [L2pro] 1. Beginner 2. Intermediate 3. Advanced 4. Unspecified 8. Instructional status [Sch] 1. Elementary school 2. Secondary school 3. University 4. Language program 5. Unspecified
Research Design	9. Type of research [R_Typ] 1. Experimental 2. Quasi-experimental 10. Interval of pretest and treatment [IntraPT] 1. Immediate 2. 1 days ~ 7 days 3. More than 1 week 4. Unspecified 11. Interval of treatment and posttest [IntraTP] 1. Immediate 2. 1 days ~ 7 days 3. 8 days ~ 14 days 4. 15 days ~ 21 days 5. More than 3 weeks 6. Unspecified 12. Interval of immediate and delayed posttest [IntraPD] 1. Immediate 2. 1 days ~ 7 days 3. 8 days ~ 14 days 4. 15 days ~ 21 days 5. More than 3 weeks 6. Unspecified
Treatment	13. Number of target of corrective feedback provided [N_target_F] 1. One 2. Two 3. Three ~ five 4. More than 5 5. Others (specify: ___) 14. Scope of feedback [Scope_F] 1. Focused 2. Unfocused 3. Others (specify:___) 15. Types of feedback [Ty_F] 1. Indirect 2. Direct 3. Both 4. Others

APPENDIX A

Continued

	16. Total session of feedback provided before posttests [Session_F] 1. One shot 2. Two 3. More than 3 17. Genres of writing tasks for treatment [Typ_task_F] 1. Journal 2. Letter writing 3. Picture-based narrative 4. Other genres 18. Target language 1. English 2. Spanish 3. Others (specify: ___) 19. Outcome measures [Out_M] 1. Accuracy rate 2. Error rate
Effect Size Data	20. Effect size based on [Typ_ES] 1. Means and standard deviation 2. *t*-value or F-value 3. Chi-square ($df = 1$) 4. Other 21. Mean and standard deviation a) Total sample size [Total_N] b) Treatment group sample size [T_N] c) Treatment group mean [T_M] d) Treatment group standard deviation [T_SD] e) Control group sample size [C_N] f) Control group mean [C_M] g) Control group deviation [C_SD]

APPENDIX B

Moderator variables

Study	Setting	School	Proficiency Level	Scope	Type	Treatment Session	TL	Genres of Writing Tasks [Practiced]
1. Semke (1980)	FL	University	Beginning	Unfocused	Direct + indirect	3	German	Journal
2. Kepner (1991)	FL	University	Intermediate	Unfocused	Direct	3	Spanish	Journal
3. Fazio (2001)	SL	Primary school	NA	Unfocused	Direct	3	French	Journal
4. Chandler (2003) Study 1	SL	University	Intermediate	Unfocused	Direct	3	English	(Autobiographical) Composition
5. Bitchener et al. (2005)	SL	Lg program	Intermediate	Focused	Direct	3	English	Letter
6. Sheen (2007)	SL	Lg program	Intermediate	Focused	Direct	2	English	Composition
7. Truscott & Hsu (2008)	FL	University	Intermediate	Unfocused	Indirect	1	English	Composition
8. Ellis et al. (2008)	FL	University	Intermediate	Focused	Direct	2	English	Composition
9. Bitchener (2008)	SL	Lg program	Intermediate	Focused	Direct	2	English	Composition
10. Bitchener & Knoch (2008)	SL	Lg program	Intermediate	Focused	Direct	1	English	Composition
11. van Beuningen et al. (2008)	SL	Secondary school	NA	Unfocused	Direct vs. indirect	2	Dutch	Letter
12. Sheen et al. (2009)	SL	Lg program	Intermediate	Focused vs. unfocused	Direct	2	English	Composition
13. Bitchener & Knoch (2009)	SL	Lg program	Intermediate	Focused	Direct	1	English	Composition
14. Bitchener & Knoch (2010b)	SL	University	Advanced	Focused	Indirect	1	English	Composition
15. Jhowry (2010)	SL	University	NA	Focused	Indirect	3	English	Composition
16. Sheen (2010)	SL	Lg program	Intermediate	Focused	Direct	2	English	Composition
17. Hartshorn et al. (2010)	SL	Lg program	Advanced	Unfocused	Indirect	3	English	Composition
18. Evans, Hartshorn, & Strong-Krause (2011)	SL	University	Advanced	Unfocused	Indirect	3	English	Composition
19. Sun (2013)	FL	University	NA (4th semester German course)	Focused vs. unfocused	Indirect + direct	3	German	Composition
20. Shintani & Ellis (2013)	SL	Lg program	Intermediate	Focused	Direct	1	English	Composition
21. Mubarak (2013)	FL	University	NA	Unfocused	Direct vs. Indirect	3	English	Various Genres of Writing

APPENDIX C

Effect Size of Individual Studies

Primary Studies	Effect Size (g)	Standard Error	95% CI Lower Limit	95% CI Upper limit
1. Semke (1980)	0.089	0.179	−0.261	0.440
2. Kepner (1991)	0.383	0.257	−0.122	0.887
3. Fazio (2001)	0.481	0.355	−0.215	1.177
4. Chandler (2003) Study 1	0.496	0.356	−0.201	1.193
5. Bitchener et al. (2005)	0.103	0.335	−0.553	0.760
6. Sheen (2007)	0.104	0.257	−0.401	0.608
7. Truscott & Hsu (2008)	0.068	0.289	−0.498	0.633
8. Ellis et al. (2008)	0.430	0.359	−0.273	1.133
9. Bitchener (2008)	1.482	0.351	0.793	2.171
10. Bitchener, & Knoch (2008)	1.375	0.257	0.872	1.878
11. van Beuningen et al. (2008)	0.888	0.388	0.127	1.648
12. Sheen et al. (2009)	0.570	0,314	−0.045	1.185
13. Bitchener, & Knoch (2010a)	0.642	0.390	−0.123	1.406
14. Bitchener & Knoch (2010b)	1.161	0.364	0.447	1.875
15. Jhowry (2010)	0.341	0.432	−0.505	1.187
17. Hartshorn et al. (2010)	0.607	0.299	0.022	1.193
18. Evans et al. (2011)	0.473	0.361	−0.235	1.181
19. Sun (2013)	0.472	0.423	−0.357	1.301
20. Shintani & Ellis (2013)	0.902	0.368	0.180	1.624
21.Mubarak (2013)	0.245	0.363	−0.466	0.956