

## **Efficacy of Written Corrective Feedback in Writing Instruction: A Meta-Analysis**

November 2020 – Volume 24, Number 3

**See Chen Lim**

National Institute of Education, Nanyang Technological University  
<lynn.freesia@gmail.com>

**Willy A Renandya**

National Institute of Education, Nanyang Technological University  
<willy.renandya@nie.edu.sg>

### **Abstract**

*This article takes a meta-analytic approach to investigate the efficacy of written corrective feedback in second language (L2) writing instruction. With its widespread practice in second language acquisition (SLA), written corrective feedback has increasingly garnered attention as much as it has continued to be a subject of considerable controversy. Aggregating findings from 35 primary studies, the study synthesizes and reviews current empirical research in the field. It seeks to complement previous meta-analyses by incorporating more recent studies and varying the inclusion criteria. Through the examination of ten moderator variables, it also aims to shed light on some factors that might mitigate the efficacy of written corrective feedback. Findings reveal a moderate overall effect, indicating that written corrective feedback has the potential to improve L2 written grammatical accuracy. In addition, direct feedback demonstrated a larger effect size than indirect feedback, though differences were not statistically significant. The study also identified learners' proficiency to be the strongest mitigator. Implications regarding how some of the key findings can be utilized to maximize the impact of written corrective feedback in L2 writing instruction are discussed.*

**Keywords:** *written corrective feedback, meta-analysis, second language acquisition, direct feedback, indirect feedback, focused feedback, unfocused feedback*

### **A Need to Examine Factors that Mitigate the Effect of Corrective Feedback**

When reviewing second language (L2) learners' writing, teachers devote substantial amounts of time responding to and correcting learners' errors. Many consider such corrective feedback a requisite to enhancing L2 learners' writing abilities (Ferris, 2010). However, despite strong conviction in its efficacy, written corrective feedback continues to be a contentious subject.

Different feedback strategies have been found to be beneficial to varying extents. Although some evidence suggests its ability to support writing accuracy in text revisions (Ferris, 1999, 2006), it is unclear if the linguistic gains in these texts can well predict accuracy improvement in subsequent writing (Sachs & Polio, 2007; Truscott & Hsu, 2008). Currently, grounded within a cognitive and psycholinguistic framework in SLA, attention seems to be steering towards written corrective feedback's potential in fostering learners' interlanguage development, examining if L2 learning is attainable when written corrective feedback is presented and acted upon (e.g., Sheen, 2007; Ellis et al., 2008; Bitchener & Knoch, 2010a). With such on-going interest and cumulation of knowledge in the field, it is timely to adopt a meta-analytic approach to synthesise and review existing empirical research. This study intends to do so in an attempt to shed light on some factors that might mitigate the efficacy of written corrective feedback.

### **Theoretical Perspectives on the Role of Corrective Feedback in SLA**

One of the theoretical claims of corrective feedback in facilitating language acquisition is the Noticing Hypothesis (Schmidt, 1990, 1995, 2000). It proposes that conscious attention and awareness of the linguistic attributes are central to language learning. Functioning as a stimulus, corrective feedback triggers learners to “notice the gap” (Schmidt & Frota, 1986) between their interlanguage output and the input of the target language (i.e., the feedback given). Subsequently, these noticing processes could stimulate destabilisation (e.g., Long, 1996; Gass, 1997; Samuda, 2001; Luchini, 2007), thus helping learners to later modify and restructure their interlanguage.

From a sociocultural perspective, Vygotsky's (1978, 1981) Sociocultural Theory suggests that language learning is mediated when learners interact with “more knowledgeable others” (1978, p. 86) who have higher language proficiency. This intervention is in the form of corrective feedback. To add value to L2 learning, such feedback has to align with the learners' zone of proximal development (ZPD). This is the region between their current and potential levels (Vygotsky, 1978, 1981). Scaffolding helps move the learner progressively towards greater understanding and independence.

From the perspective of DeKeyser's (2003, 2007a, 2007b) Skill Acquisition Theory, corrective feedback stimulates learners' declarative knowledge, transforming it into procedural knowledge (Bitchener, 2012). Eventually, the learner progresses from controlled to automatic processing with less attention, faster pace and greater accuracy. It also provides explicit knowledge and prevents incorrect information from becoming proceduralised and getting executed in an automatic manner (Polio, 2012). While emphasising that meaningful and ample practice is necessary to achieve automaticity, DeKeyser (2007b) also advises that further investigation is needed to ascertain the amount and nature of feedback that is useful during practice.

Another supporting theoretical notion, Ellis's (2010) componential framework for corrective feedback, explains how learners' individual difference factors (e.g. age, motivation, learning style, and beliefs) might interact with contextual factors (e.g. learning setting) to mediate between the oral and written corrective feedback learners receive and their engagement with it. Learner engagement is examined from three angles – cognitive (how learners attend to corrective feedback), behavioural (learners' uptake or revision due to corrective feedback) and attitudinal (learners' attitudes to corrective feedback such as aversion or anxiety). The framework attempts to identify the variables that corrective feedback studies have addressed

thus far, and suggests areas that future studies should consider such as the influence of individual difference factors.

While many researchers support the use of corrective feedback in L2 instruction, there are objections against it. For instance, Truscott (1996) argues that this form of ‘pseudo-learning’ could only lead to superficial, explicit knowledge, and not the implicit knowledge that is required for language acquisition. Concerned with the developmental readiness of learners, he contends that as long as a learner is not prepared to acquire the written corrective feedback, intake may not transpire. This corresponds with Krashen’s (1981; 1982; 1985) Natural Order Hypothesis, which postulates that learners acquire different grammatical structures in a somewhat predefined sequence and not in the order determined by the teacher or the syllabus. It is also echoed by Pienemann’s (1989) Learnability Hypothesis which stipulates that only when learners demonstrate developmental readiness can they acquire the structures. Truscott (1996, 2004) also argues that the great amount of effort and time spent on managing feedback and corrections divert teachers’ and learners’ attention away from important tasks (such as extra writing practices) which are more likely to promote acquisition. In addition, it triggers a high affective filter in learners, raising their anxiety levels and lowering their self-esteem. Learner distress may also cause them to avoid erroneous structures in new texts and hinder them from being more explorative with the language, resulting in simplified writing (e.g., Krashen, 1982). Finally, with many teachers in the EFL setting having varied levels of proficiency, the capacity of L2 teachers to provide adequate and consistent feedback has been called into question.

### **Empirical Evidence on Corrective Feedback in SLA**

Earlier studies on corrective feedback (e.g., Semke, 1984; Robb et al., 1986; Kepner, 1991; Sheppard, 1992) have been criticized due to design flaws and poor outcome measures such as the use of ‘content-comments’ group (instead of a strictly no-feedback control group) for comparison. Current studies have attempted to be more tightly controlled in this aspect.

In general, the two most commonly studied dichotomies are direct-indirect feedback and focused-unfocused feedback. Despite multiple studies to investigate the relative merits of direct (correct linguistic forms are provided above the error directly) and indirect (presence of an error is indicated without explicit correction) written corrective feedback, no strong conclusions have been reached. Ferris (2010) speculates that the conflicting results could be due to their disparities in research focus. While direct feedback can efficiently trace evidence for the mastery of specific structures, indirect feedback can better evaluate how successful a potential strategy for metacognitive skills is. Many researchers (e.g., Ellis, 2009; Ferris & Roberts, 2001) believe that the efficacy of either feedback is influenced by learner proficiency. Ellis (2009), for instance, claims that beginners may benefit more from direct feedback as they still need strong support to broaden their linguistic repertoire. Conversely, advanced learners have the ability to discern errors independently, making indirect feedback a more appropriate choice. Likewise, Liu (2008) affirms that since learners with lower proficiency find it challenging to perform self-corrections, direct feedback may be a more effective strategy.

Similarly, the debate between focused (targeting only one or a few specific types of linguistic error) and unfocused feedback (correcting a broad range of error categories) has no clear winner at this point. Many studies focused on only a few error types, particularly the English definite and indefinite articles. As Xu (2009) has pointed out, it is not feasible to predict if written corrective feedback can successfully treat other linguistic errors based on the positive effects

of this structure alone. Storch (2010) echoes the same sentiments that highly-focused feedback studies are ecologically less significant, despite being more methodologically rigorous. Liu and Brown (2015) also note that increasingly, studies are bridging theory and practice by focusing on a broader range of errors with more authentic classroom treatments. More recently, Li & Vuono (2019) pointed out that a missing dimension in the literature is whether feedback is comprehensive (addressing all errors) or selective (targeting selected errors). They proposed that for instance, if researchers use focused feedback to correct errors pertaining to one linguistic structure, the decision of correcting all or selected errors within that particular structure will still have to be made clear.

### **Meta-analyses related to the Effectiveness of Corrective Feedback in SLA**

Ever since Norris and Ortega (2000) published their seminal research on the efficacy of L2 instruction in general, meta-analysis has been recognised for its important role in the field of SLA. One such instruction type was corrective feedback. Following this, several other meta-analyses emerged (see [Appendix A](#)). While existing meta-analyses have looked into the efficacy of corrective feedback, the foci and selection criteria differ. Furthermore, with an expanding database on corrective feedback, an update is necessary in order to gain fresh insights, particularly into the realms of written corrective feedback. Thus, the goal of the present meta-analysis is to build on and complement existing research by examining written corrective feedback as the sole construct while probing into more mitigating variables. It seeks to address the following questions:

- RQ1.** What is the overall effect of written corrective feedback on improving L2 written accuracy?
- RQ2.** Which type of written corrective feedback is more effective?
- RQ3.** What factors might mitigate the efficacy of written corrective feedback?

## **Method**

### **Identifying Primary Studies**

To develop a decent database of primary studies, the subsequent steps were taken. First, three frequently-used online databases in the field of education and applied linguistics were accessed. These are Education Resources Information Center (ERIC), LLBA (Linguistics and Language Behaviour Abstract), and ProQuest. The following key words and combination of key words were used: *written corrective feedback, direct feedback, indirect feedback, error correction, second language acquisition/learning*.

Second, manual and electronic searches were conducted for current and past issues of several widely cited journals in applied linguistics and SLA. These journals were determined by examining the reference lists of previous review papers (e.g., Storch, 2010; van Beuningen, 2010) as well as meta-analyses on corrective feedback. Such journals include *Language Learning, Language Teaching, Studies in Second Language Acquisition, Applied Linguistics, The Modern Language Journal, TESOL Quarterly, Language Teaching Research, System, The Canadian Modern Language Review, International Review of Applied Linguistics, English Language Teaching, International Journal of English Studies, ELT Journal, and Journal of Second Language Writing*.

Third, the reference sections of book chapters related to written corrective feedback (e.g., Bitchener & Ferris, 2012), published meta-analyses, narrative reviews (e.g., Bitchener, 2012; Biber, 2011; Li, 2010), state-of-the-art articles (e.g., Nassaji, 2016; Ellis & Sheen, 2006) and edited books related to corrective feedback were checked to locate potential sources of primary studies.

Fourth, Google Scholar was used to verify studies or search for additional ones. To minimize availability bias (“file-drawer” problem), the current study also included Ph.D. and Master’s Dissertations. Upon identification of the prospective studies, the article abstracts were meticulously screened using a predetermined list of inclusion/exclusion criteria. A sample of 38 primary studies was thus retrieved.

### **Inclusion/Exclusion Criteria**

To be included in the current meta-analysis, studies had to pass the following screening criteria:

1. The study had to be an experimental or quasi-experimental research which used a control group (without feedback) as comparison. Such a rationale stems from many L2 researchers’ view that the effects of learning after treatment could be measured by comparing the gains of the control and the treatment groups (e.g., Ferris, 1999, 2004; Truscott, 2004). Therefore, studies without control groups (e.g., Liu, 2008; Suzuki et al., 2018) were excluded.
2. The study had to examine L2 written corrective feedback provided by a teacher instead of by peers or via a computer.
3. Learners’ written grammatical accuracy is measured, rather than their beliefs, preferences, attitudes or perceptions.
4. The study had to be published in English.
5. The study had to be published in 2001 or later. For the current study, the cut-off date was set at June 2019.

In addition, the following points had to be noted:

1. A number of studies (e.g., Ferris et al., 2002; Lee, 2004) employed qualitative (and often anecdotal) methods. As these studies did not provide statistics from experiments or quasi-experiments, they were not considered in the current meta-analysis.
2. Studies which did not report sufficient statistics required for computation of effect sizes were also excluded (e.g., Hosseiny, 2014; Ferris et al., 2013).
3. When multiple studies report the same experimental result, such as Bitchener and Knoch (2009a) and Bitchener and Knoch (2010a), only one was used in the meta-analysis. The latter study was included in this case.
4. When studies employ a design such that it was impossible to isolate the effects of specific feedback from other types of treatment, they were excluded. For instance, in Bitchener & Knoch (2009b), the treatment group comprised different subgroups of corrective feedback (direct, oral and written meta-linguistic explanation), thus, the effects of direct feedback could not be disentangled from other instruction types. As such, this study was omitted.

### **Data Coding**

The coding scheme for the 38 studies (33 published studies and 5 unpublished Ph.D. and Master’s Dissertations) was devised with reference to some of the common variables considered in education and applied linguistics meta-analytic features. Variables and descriptors of the coding protocol are described in Table 1.

**Table 1. Coding Scheme**

Main Category	Variables	Descriptors
Study Characteristics (8)	Study ID	–
	Author	–
	Learners' age	<ul style="list-style-type: none"> <li>• Mean age of participants</li> </ul>
	Publication year	<ul style="list-style-type: none"> <li>• Before 2008</li> <li>• From 2008</li> </ul>
	Publication type	<ul style="list-style-type: none"> <li>• Published article</li> <li>• Dissertation</li> </ul>
	Research setting	<ul style="list-style-type: none"> <li>• Foreign Language</li> <li>• Second Language</li> </ul>
	Instructional status	<ul style="list-style-type: none"> <li>• Primary School</li> <li>• Secondary School</li> <li>• Junior College</li> <li>• Language programme</li> <li>• University</li> </ul>
	Learners' proficiency level	<ul style="list-style-type: none"> <li>• Low/Low Intermediate</li> <li>• Intermediate</li> <li>• High Intermediate/Advanced</li> </ul>
Research (4)	Design	
	Type of research	<ul style="list-style-type: none"> <li>• Experimental</li> <li>• Quasi-experimental</li> </ul>
	Interval between pretest and treatment	<ul style="list-style-type: none"> <li>• Immediate</li> <li>• 1 – 7 days</li> <li>• More than 1 week</li> </ul>
	Interval between treatment and posttest	<ul style="list-style-type: none"> <li>• Immediate</li> <li>• 1 – 7 days</li> <li>• 8 – 14 days</li> <li>• More than 2 weeks</li> </ul>
Treatment (8)	Number of target features	<ul style="list-style-type: none"> <li>• One</li> <li>• Two</li> <li>• Three or more</li> </ul>
	Length of treatment	<ul style="list-style-type: none"> <li>• Short (<math>\leq 45</math> min)</li> <li>• Medium (50 – 100 min)</li> <li>• Long (<math>&gt; 100</math> min)</li> </ul>
	Total number of treatment sessions	<ul style="list-style-type: none"> <li>• One-shot</li> <li>• 2 or 3</li> <li>• <math>&gt; 3</math></li> </ul>
	Number of treatment sessions before immediate posttest	<ul style="list-style-type: none"> <li>• 1</li> <li>• 2 or 3</li> <li>• <math>&gt; 3</math></li> </ul>
	Scope of feedback	<ul style="list-style-type: none"> <li>• Focused</li> <li>• Unfocused</li> </ul>
	Type of feedback	<ul style="list-style-type: none"> <li>• Direct</li> <li>• Indirect</li> </ul>
	Target language	<ul style="list-style-type: none"> <li>• English</li> <li>• Non-English</li> </ul>
	Outcome measure	<ul style="list-style-type: none"> <li>• Written grammatical accuracy</li> </ul>
	Effect (7)	Size
Total sample size		–
Treatment group size		–
Treatment group mean		–
Treatment group standard deviation		–
Control group size		–
Control group mean		–
Control group standard deviation	–	

## **Moderator Variables**

Ten variables were identified as moderator variables for the current study (see [Appendix B](#)). These moderator variables were selected based on recommendations from previous studies. Firstly, past meta-analyses (e.g., Russell & Spada, 2006; Li, 2010) have suggested that settings and proficiency of learners have an impact on how effective written corrective feedback is. Secondly, some researchers have also pointed out that individual learners respond differently to the scope and type of feedback (e.g., Ferris, 2002; Ellis, 2009). Thus, findings derived from the investigation of these effect sizes may inform teachers on the kinds of feedback to deploy in the classroom. Additionally, it would be worth noting how instructional status and the publication type affect its effectiveness, as well as how more recent studies compare to older ones. A new category was introduced in the present study – the number of treatments before immediate posttest. This was deemed necessary because the overall effect was calculated using just the immediate posttest measures. Thus, inclusion of such a category might give a more comprehensive picture of the analysis.

Variables related to Study Characteristics were coded in the following manner. Setting was coded into foreign language or second language. In the former, students from non-English speaking backgrounds learn the language in the context where English is not widely used for communication or as a medium of instruction. In the latter, English is the predominant language of communication. Instructional status was coded into primary school, secondary school, junior college, language programme or university, as reported in the primary studies. Due to different descriptors used by the primary researchers, coding for learners' proficiency was a little more complicated. To avoid the classifications from being too finely divided, three categories were decided for this study. Studies which reported learners' proficiency as low, low intermediate, and pre-intermediate were subsumed under the low/low intermediate category. Studies which reported learners' proficiency as high, high intermediate, post intermediate, upper intermediate and advanced were categorised as high intermediate/advanced. The third category 'intermediate' was more clear-cut – it was set aside for those that reported proficiency as intermediate. Finally, publication type was classified as published article or dissertation, and publication year was coded into before 2008 and from 2008.

Variables related to Treatment were coded as follows. Scope of feedback was coded into focused and unfocused. Type of feedback was coded into direct and indirect. Target language was coded into English and non-English (which comprised a French, a German, and two Dutch studies). Total number of treatment (feedback) sessions was coded into one shot, 2 or 3, or more than 3. Finally, the number of treatments before immediate posttest was coded as 1, 2 or 3, or more than 3. To ensure reliability in coding, the primary studies were coded five times. Subsequently, a second coder coded 6 studies (including 2 dissertations and 4 published articles picked randomly). The interrater agreement rate was 97% and differences were resolved through discussion.

## **Data Analysis**

All the analyses for the present study were performed using a professional meta-analysis program known as Comprehensive Meta-Analysis (CMA) (Borenstein et al., 2005). The decision to use this software was also based on its versatility, reliability, and positive feedback. Hunter and Schmidt (2004, p. 466) described it as an almost "all-purpose meta-analysis

program". Littell, Corcoran and Pillai (2008, p. 146) lauded it as "probably the most sophisticated stand-alone package for meta-analysis". Since its development, the program has been adopted in many meta-analyses across different academic disciplines.

Effect sizes for the study were calculated using the outcome measures of written grammatical accuracy. All measures were given as continuous variables, primarily in the form of error and accuracy rate of learners' writing. The means, standard deviations and sample sizes of the treatment and control groups in each study were teased out and used for the computation of the effect sizes. These are measured in terms of Hedges's  $g$  (a conservative version of Cohen's  $d$ ) due to its ability to correct for biases as a result of small sample sizes (Lipsey & Wilson, 2001).

With the underlying assumption that a true effect size varies across studies, the random effects model was adopted to yield an average effect size. This is plausible because the participants and interventions in these studies (such as setting, age and language proficiency) would have differed in ways that would have influenced the results, thus a common effect size should not be assumed.

As cautioned by Lipsey and Wilson (2001) that using more than one effective size in a meta-analysis would result in an inflated sample size and dubious statistical results, the principle of "one study, one effect size" was adopted and consistently adhered to in the study. Yet, in reality, numerous studies incorporated multiple treatments, subgroups and outcomes measures, thus contributing more than one effect size. To overcome this predicament, multiple effect sizes related to a single construct within a study were averaged. For instance, Sheen et al. (2009) investigated direct focused vs. direct unfocused feedback. Being subsets of the larger category of direct corrective feedback, these two effect sizes were averaged in the final analysis for the overall effect. Likewise, Shintani, Ellis and Suzuki (2014) adopted multiple outcome measures of the accuracy rates of the indefinite article and hypothetical conditional. Since these belong to a broader construct of grammatical accuracy, effect sizes from the multiple measures were averaged. Finally, as noted earlier, only measures from the immediate posttests were used to calculate for overall effect. This is because not all studies conduct delayed posttests.

Upon obtaining an estimate of the overall effect size of written corrective feedback,  $Q$  statistics were employed to check the heterogeneity across effect sizes. A significant  $Q$  would indicate that the effect sizes are not homogeneous. Disparities in variables such as the type and scope of written corrective feedback, outcomes and measures, are likely to give rise to heterogeneity. Rejecting the null hypothesis of homogeneity justifies the test for moderator effects. This will then allow for identification of potential variables which may mitigate the effects of written corrective feedback. Despite indicating the statistical significance of heterogeneity, the  $Q$  statistic and its  $p$ -value do not reveal the precise amount of dispersion. Further computations of  $T^2$  and  $I^2$  statistics were thus performed to elicit the degree of heterogeneity.  $T^2$  reflects the amount of true heterogeneity while  $I^2$  shows the "proportion of observed dispersion that is real" (Borenstein et al., 2011, p. 125).

Subsequently, an analysis of moderator variables was performed to investigate the sources of heterogeneity across the effect sizes. For each moderator variable identified for this meta-analysis, only the studies that provided such information were considered. For instance, if the learners' proficiency level was not reported in a study (e.g., Murabak, 2013), the effect size data from this particular study was excluded from the analysis of the moderator variable on 'Proficiency Level.' However, it was still included in the analysis of other moderators, as well

as in the calculation of the overall mean effect size. To assess the extent to which potential moderator variables account for the variance in the effects of written corrective feedback,  $Q$  Between ( $Q_b$ ) tests were also performed.  $Q_b$  values thus indicate whether effect sizes vary significantly across the moderator variables.

Altogether, 38 studies were found for the preliminary analysis, among which 33 were published.

### **Ethics**

The ethical considerations of this meta-analysis relate to the process of reporting and publishing results. First, conducting such an analysis involves gathering, summarising and integrating huge amounts of data, thus extra vigilance was exercised in extracting and analysing data accurately. Second, the inclusion and exclusion criteria were stated explicitly and applied consistently. Third, tests for publication bias were also established.

## **Results**

### **Preliminary Analysis and Outliers**

Altogether, 38 studies were surfaced for the preliminary analysis, among which 33 were published articles, 5 were unpublished dissertations (4 Ph.D. and 1 Master's). Based on this sample, written corrective feedback appears to have a large effect on the grammatical accuracy of L2 students' writing (Hedges's  $g = 0.72$ ,  $SE = 0.106$ ,  $CI = 0.516 \sim 0.930$ ,  $p < 0.0001$ ). There was also significant heterogeneity across the effect sizes of the studies ( $Q = 150.244$ ,  $p < 0.0001$ ). The variance of true effects  $T^2$  was 0.314 and  $I^2$  was 75.373. This implied that 75.38% of the observed heterogeneity could be due to real differences between studies, and not random error. These results warranted the subsequent analysis of moderator variables to justify the variance in effect sizes across the studies. However, having a relatively small sample size of 38 (which is often so for SLA meta-analyses), extreme values in the analysis could impact the results substantially. Thus, it was crucial to check for outliers. To do so,  $Z$  scores (the standardised estimates of effect sizes) were examined. The  $Z$  scores of three studies – Sarvestani (2015), Samiei et al. (2016) and Nemati et al. (2019) indicated that each of their effect sizes was at least five standard deviations above the overall effect size. When these outliers were omitted, the overall effect size was reduced to  $g = 0.59$  ( $CI = 0.423 \sim 0.755$ ). Therefore, to ensure robustness of results, these three studies were consequently excluded from all analyses.

### **Overall Effect of Written Corrective Feedback**

The final sample of 35 studies was used to address the first research question on the overall effect of written corrective feedback. An overall effect size of  $g = 0.59$  ( $SE = 0.085$ ,  $CI = 0.423 \sim 0.755$  and  $p < 0.0001$ ) was obtained, indicating that there was still significant heterogeneity across the effect sizes ( $Q = 83.111$ ,  $p < 0.0001$ ). The variance of true effects  $T^2$  was 0.144 and  $I^2$  was 59.091. Table 2 summarises the effect sizes of individual studies. Most studies reported positive effects for written corrective feedback while a few (indicated by a negative sign) found otherwise.

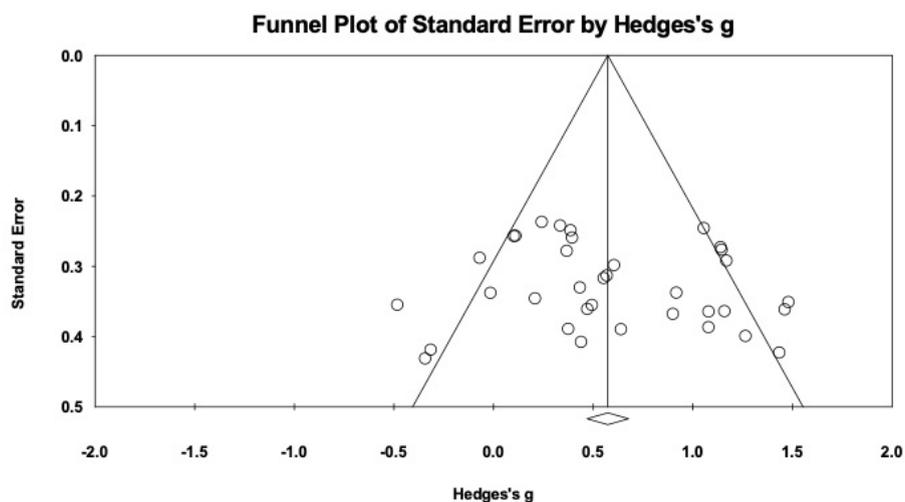
**Table 2. Effect Size of Individual Studies.**

No.	Primary Studies	Effect Size (g)	Standard Error	95% CI Lower Limit	95% CI Upper Limit
1	Fazio (2001)	-0.481	0.355	-1.177	0.215
2	Ferris & Roberts (2001)	0.435	0.331	-0.213	1.084
3	Chandler (2003)	0.496	0.356	-0.201	1.193
4	Bitchener (2005)	-0.013	0.338	-0.677	-0.650
5	Sheen (2007)	0.104	0.257	-0.401	0.608
6	Truscott & Hsu (2008)	-0.068	0.289	-0.633	0.498
7	Bitchener & Knoch (2008)	1.057	0.246	0.575	1.540
8	Bitchener (2008)	1.482	0.351	0.793	2.171
9	Ellis et al. (2008)	0.441	0.408	-0.359	1.241
10	Beuningen et al. (2008)	0.210	0.346	-0.468	0.889
11	Sheen et al. (2009)	0.570	0.314	-0.045	1.185
12	Bitchener & Knoch (2010a)	0.642	0.390	-0.123	1.406
13	Bitchener & Knoch (2010b)	1.161	0.364	0.447	1.875
14	Hartshorn et al. (2010)	0.607	0.299	0.021	1.192
15	Sheen (2010)	0.396	0.260	-0.113	0.905
16	Jhowry (2010)	-0.341	0.432	-1.187	0.505
17	Evans et al. (2011)	0.473	0.361	-0.235	1.181
18	Beuningen et al. (2012)	0.337	0.243	-0.139	0.812
19	Farrokhi & Sattarpour (2012)	1.464	0.362	0.754	2.173
20	Ahmadi et al. (2012)	0.555	0.317	-0.067	1.177
21	Shintani & Ellis (2013)	0.902	0.368	0.180	1.624
22	Murabak (2013)	1.081	0.387	0.322	1.840
23	Karim (2013)	0.377	0.390	-0.387	1.140
24	Kassim & Ng (2013)	1.147	0.277	0.604	1.690
25	Sun (2013)	-0.313	0.419	-1.135	0.508
26	Jiang & Xiao (2014)	1.141	0.273	0.606	1.677
27	Shintani, Ellis & Suzuki (2014)	1.171	0.293	0.597	1.744
28	Hartshorn & Evans (2015)	1.436	0.423	0.607	2.266
29	Stefanou & Révész (2015)	0.112	0.257	-0.392	0.616
30	Frear & Chiu (2015)	1.267	0.400	0.483	2.050
31	Rezazadeh (2015)	0.388	0.249	-0.101	0.877
32	Hosseini (2015)	0.919	0.338	0.257	1.582
33	Khanlarzadeh & Nemati (2016)	1.081	0.365	0.366	1.797
34	Wang (2017)	0.244	0.237	-0.221	0.709
35	Guo & Barrot (2019)	0.369	0.279	-0.177	0.915

As mentioned previously, delayed posttests were left out from the overall effect analysis with the purpose of preserving the independence of effect sizes. Of the 35 studies, 23 incorporated at least one delayed posttest, and these delayed effect sizes were analysed. For studies which reported more than one delayed posttest, the first one was used for analysis. On the whole, the delayed posttests were conducted at least 2 weeks after the corrective feedback treatment. Their mean effect size was  $g = 0.569$  ( $SE = 0.118$ ,  $CI = 0.337 \sim 0.801$  and  $p < 0.0001$ ). This suggests moderate effects for learning retention. Nevertheless, due to a small sample size, these results need to be interpreted cautiously.

### Publication Bias

In order to check for any publication bias (i.e., whether the studies retrieved were inclined to be those with significant findings), a funnel plot of effect sizes across 35 studies was examined. Estimates of the effect sizes of studies with large sample sizes appear towards the apex of the funnel, while those with smaller sample sizes cluster around the bottom (Duval & Tweedie, 2000). If the funnel plot appeared to be roughly symmetric, it is unlikely that publication bias was present, implying that the meta-analysis had included all the relevant studies (Borenstein et al., 2011). On the contrary, missing studies in the dataset would result in an asymmetric funnel shape. As illustrated in Fig. 1, studies captured in the present meta-analysis are equally dispersed on either side of the overall effect, thus denoting that the sample is free from publication bias.



**Figure 1. Funnel Plot of 35 Studies.**

To further verify the findings, a trim-and-fill (Duval, 2005) analysis was also performed to locate any missing studies that may have been ignored due to publication bias. Omitted ones (if any) would have to be included in the analysis and an overall effect size would have to be re-imputed. This analysis showed that the re-imputed estimate was the same and no value of the overall effect size was found to be missing. The random effects model revealed an overall effect size of  $g = 0.589$  ( $CI = 0.423 \sim 0.755$ ). Thus, findings concur with those of the funnel plot. Additionally, a fail-safe  $N$  was calculated. Results suggested that 1,030 studies would have to be neglected to invalidate a significant effect size result. This number surpassed the criterion number of  $5k + 10 = 180$  where  $k = 35$  studies (Rosenthal, 1991). Therefore, the fail-safe  $N$  results further confirmed the absence of publication bias in this meta-analysis.

## Moderator Analyses

To establish if learner characteristics and methodological features influence the effectiveness of corrective feedback, separate analyses were performed by applying effect sizes related to only the immediate posttests. This addressed the third research question about the factors mitigating the efficacy of written corrective feedback. Random effects models were predominantly employed. However, as recommended by Borenstein et al. (2011), when the sample size of a subgroup is small, it may not be possible to estimate the between studies variance (tau-squared) with precision. A fixed effects model would then be viable. Thus, in this study, if the subgroup sample size was less than five, a fixed effects analysis was also performed to give a more comprehensive picture. A summary of the moderator variables and their respective effect sizes is shown in Table 3.

The study setting was first analysed as a moderator variable. Written corrective feedback was found to have a greater effect when provided in a foreign language setting ( $g = 0.703$ ) than in a second language one ( $g = 0.478$ ). However, the differences were not significant ( $Q = 1.776$ ,  $p = 0.183$ ).

In terms of instructional status, the only study conducted in a junior college exhibited the largest mean effect size ( $g = 1.141$ ), whereas that in the primary school yielded a negative effect ( $g = -0.481$ ). These differences were not significant under the random effects model ( $Q = 9.330$ ,  $p = 0.053$ ). As there was only one sample each in the primary school and junior college category, and three samples in the secondary school category, the fixed effects model was also employed. This analysis yielded significant differences among the five categories ( $Q = 19.534$ ,  $p = 0.001$ ).

With regard to proficiency of the learners, significant differences were observed among the three levels ( $Q = 12.965$ ,  $p = 0.002$ ). Learners with low/low intermediate proficiency seemed to gain most from written corrective feedback ( $g = 0.982$ ). This is followed by high intermediate/advanced learners ( $g = 0.696$ ). Intermediate learners, however, did not seem to benefit as much ( $g = 0.364$ ).

The types of feedback did not produce any significant differences between them ( $Q = 0.322$ ,  $p = 0.570$ ), though the effect size for direct feedback ( $g = 0.761$ ) was higher than indirect feedback ( $g = 0.625$ ).

Similarly, analysis of the scope of feedback did not present any significant differences between focused and unfocused feedback ( $Q = 1.050$ ,  $p = 0.305$ ). Nonetheless, the effect size for the former ( $g = 0.628$ ) was higher than its counterpart ( $g = 0.445$ ).

In the total number of treatments category, studies which provided one shot treatment ( $g = 0.641$ ) outperformed those offering multiple sessions. Studies that provided two to three treatment sessions yielded an effect size of  $g = 0.575$  while those that provided more than three sessions obtained  $g = 0.506$ . However, these differences were not statistically significant ( $Q = 0.415$ ,  $p = 0.813$ ).

Likewise, the number of treatments before the immediate posttest showed that a single treatment seemed superior ( $g = 0.657$ ), though again, the differences between them were not statistically significant ( $Q = 1.537$ ,  $p = 0.464$ ).

**Table 3. Summary of Moderator Variables.**

No.	Moderator Variable	<i>k</i>	Effect Size ( <i>g</i> )	95% <i>CI</i> Lower Limit	95% <i>CI</i> Upper Limit	Qb Value (p-value)
1	<b>Setting</b>					1.776 (0.183)
	Foreign Language	17	0.703	0.467	0.939	
	Second Language	18	0.478	0.245	0.710	
2	<b>Proficiency</b>					12.965 (0.002)
	Low/Low Intermediate	10	0.982	0.740	1.223	
	Intermediate	10	0.364	0.130	0.599	
	High Intermediate/Advanced	10	0.696	0.455	0.938	
3	<b>Instructional Status</b>					9.330 (0.053) / 19.534 (0.001)*
	Primary School	1	-0.481 / -0.481*	-1.442 / -1.177*	0.480 / 0.215*	
	Secondary School	3	0.222 / 0.227*	-0.274 / -0.082*	0.728 / 0.535*	
	Junior College	1	1.141 / 1.141*	0.290 / 0.606*	1.993 / 1.677*	
	Language Programme	12	0.714 / 0.687*	0.449 / 0.508*	0.980 / 0.865*	
	University	18	0.591 / 0.580*	0.371 / 0.430*	0.812 / 0.730*	
4	<b>Type of Feedback</b>					0.322 (0.570)
	Direct	16	0.761	0.514	1.007	
	Indirect	7	0.625	0.225	1.024	
5	<b>Scope of Feedback</b>					1.050 (0.305)
	Focused	16	0.628	0.398	0.857	
	Unfocused	13	0.445	0.182	0.708	
6	<b>Total no. of Treatments</b>					0.415 (0.813)
	1	16	0.641	0.396	0.886	
	2 to 3	9	0.575	0.239	0.911	
	More than 3	10	0.506	0.170	0.842	
7	<b>No. of Treatments before Immediate Posttest</b>					1.537 (0.464)
	1	20	0.657	0.437	0.876	
	2 to 3	6	0.367	-0.034	0.769	
	More than 3	9	0.585	0.237	0.933	
8	<b>Target Language</b>					7.138 (0.008) / 12.071 (0.001)*
	English	31	0.663 / 0.639*	0.497 / 0.528*	0.828 / 0.749*	
	Non-English	4	0.014 / 0.048*	-0.482 / -0.266*	0.454 / 0.362*	
9	<b>Publication Type</b>					2.069 (0.150)
	Dissertation	5	0.271	-0.192	0.734	
	Published Article	30	0.635	0.458	0.812	
10	<b>Publication Year</b>					5.974 (0.015)
	Before 2008	5	0.111	-0.302	0.524	
	From 2008	30	0.667	0.498	0.836	

\*data from fixed effects analysis, *k* = number of studies, *CI* = confidence interval

Under the random effects model, studies with English as the target language yielded a significantly higher effect ( $g = 0.663$ ,  $p = 0.008$ ) than those with non-English ( $g = -0.014$ ). As the non-English category only had a small sample size of four, the fixed effects model was also analysed. Likewise, English yielded a higher effect ( $g = 0.639$ ) than non-English ( $g = 0.048$ ), and at  $p = 0.001$ , this difference was also statistically significant.

In terms of the type of publication, the 30 published studies had a larger effect ( $g = 0.635$ ) than the 5 dissertations ( $g = 0.271$ ). However, at  $p = 0.150$ , differences between them were not statistically significant.

On the contrary, significant differences ( $Q = 5.974$ , and  $p = 0.015$ ) were found between studies published from 2008 and those published before that. The earlier studies generated a smaller effect size ( $g = 0.111$ ) than the more recent ones ( $g = 0.667$ ).

## Discussion

To address the first question about the overall effect of written corrective feedback, the study generated an overall effect size of  $g = 0.59$ . Based on benchmarks suggested by Cohen (1988), the effect size is low if  $g$  is below 0.2, medium if it is around 0.5, and large if it is greater than 0.8. However, according to Thompson (2007), these values are arbitrary and should not be interpreted rigidly. Lipsey and Wilson (2001) also cautioned that these values do not consider the context of the specific research domain. In recent years, Oswald and Plonsky (2010) proposed an SLA-specific benchmark, interpreting 0.40 to be small, 0.70 to be medium, and 1.0 to be large. Using this yardstick, the overall effect size of  $g = 0.59$  found in this meta-analysis is close to moderate. This result stands in stark contrast to Truscott's (2007) findings of  $d = -0.155$  for controlled studies. Such a discrepancy could be due to the differences in the primary studies included for both studies. While Truscott (2007) chose only published studies based on narrative reviews with 2006 as the cut-off year, the current study, incorporated both published and unpublished studies from 2001 to 2019. Nevertheless, this result is comparable to other meta-analyses investigating written corrective feedback. For instance, Russell and Spada (2006) obtained a considerable effect size of  $d = 1.31$  for written corrective feedback, while Biber et al. (2011) and Kang and Han (2015) reported a moderate effect size of  $d = 0.4$  and  $g = 0.54$  respectively. The variation in effect size from these three meta-analyses could be attributed to some differences in the inclusion criteria due to slightly different foci of each study. Despite some of these variations, findings from this meta-analysis generally support most of the other recent meta-analyses, thus complementing the efficacy of written corrective feedback.

The second research question addressed which type of written corrective feedback is more effective, in particular, the direct or the indirect feedback. Though direct feedback seemed to have a large impact, statistically, the analysis did not reveal a distinct difference between the two. A possible explanation could be that the type of feedback might operate in tandem with other factors, such as the length of treatment and proficiency of learners. Thus, their effects are not clearly distinguishable. Indeed, as highlighted earlier, many researchers (e.g., Ellis, 2009; Ferris, 2002; Liu, 2008) believe in the importance of taking learners' proficiency into consideration when providing feedback. They suggest that direct feedback contributes more to the learning of those with lower proficiency levels since these learners may not have developed adequate linguistic knowledge to correct their own errors. To date, there are no empirical

studies which focus solely on how proficiency level relates to the effects of different feedback types. Such a thought-provoking issue certainly merits further investigation.

Some researchers (e.g., Ferris & Robert, 2001; Ferris, 2003) also advocate the efficacy of indirect feedback in the long term, since it requires learners to be engaged in more sophisticated language processing. Interestingly, when an analysis of the 23 studies (which reported at least one delayed posttest) was performed, it was found that under the random effects model, indirect feedback yielded an effect size of  $g = 0.998$  while direct feedback was substantially lower at  $g = 0.531$ . Fixed effects analysis yielded similar findings, with indirect feedback at  $g = 0.979$ , and direct feedback at  $g = 0.468$ . It should be noted, however, that this difference is not significant, at  $p = 0.297$  (random) and  $p = 0.067$  (fixed). Besides, the small sample size of only two indirect studies might render it difficult to make conclusive claims. Nevertheless, the stark difference calls for more attention in this area.

The third research question explores the factors that might mitigate the efficacy of written corrective feedback. Among the ten moderator variables analysed, several noteworthy findings are surfaced.

First, learners' proficiency emerged as the strongest mitigator of effect size among all the variables. This discovery coincides with Kang and Han's (2015) findings. However, in their study, larger effect sizes were found with increasing proficiency levels. But in the present study, learners with low/low intermediate proficiency seemed to reap more benefits from written corrective feedback, followed by learners with high intermediate/advanced proficiency. Comparatively, intermediate learners do not seem to gain as much. The varied findings in both meta-analyses could be attributed to the different interpretations of proficiency levels. For instance, in Shintani and Ellis's (2013) study, participants are reported as low intermediate, and while Kang and Han (2015) coded it as intermediate, the present study coded it as low/low intermediate. At this point, it must be highlighted that the decision of the primary researchers on their participants' proficiency levels were arbitrary and highly specific to their respective contexts. It is no wonder Li (2010) cited this as a rationale for not including learners' proficiency as a variable in his moderator analysis. Nevertheless, the important finding of this variable as a strong mitigator reiterates the importance of taking developmental readiness into consideration when providing feedback (Truscott, 1996; Pienemann, 1998). Additionally, it would also be useful to study how feedback could be provided in order to advance the progress of learners with intermediate proficiency levels. To get around the problem of subjectivity in determining learners' proficiency levels, future researchers could explore the possibility of adopting established measures to describe learners' proficiency standards. For instance, tests such as *Common European Framework of Reference for Languages* (CEFR), International English Language Testing System (IELTS) or Test of English as a Foreign Language (TOEFL) could be used.

In terms of setting, results from the analysis seem to suggest that learners in foreign language contexts reap more benefits from written corrective feedback than those in second language contexts. This seems at odds with Kang and Han's (2015) findings which showed otherwise. The discrepancy could be because of the differences in the sample sizes. Kang and Han (2015) had a smaller and not as balanced representation in their groups (15 for second language setting and six for the foreign language setting). In contrast, the current study featured 18 and 17 in the respective settings. Nevertheless, results of this analysis still concur with findings from several other studies on oral corrective feedback. For instance, Sheen (2004) observed that

learners in foreign language settings paid more attention to feedback when interacting. They could also modify their output accordingly as compared to those in the second language setting. Similarly, Li (2010) also reported higher effect sizes for the foreign language setting than for the second language setting. A possible reason for this phenomenon could be that learners in foreign language settings are more positive and responsive towards error correction (Loewen et al., 2009), thus, they are more likely to incorporate the given feedback. Another conjecture is that the instructional dynamics in foreign language contexts might likely increase the efficacy of corrective feedback. Additionally, in Liu's (2007) survey involving 800 teachers of English globally, he realised that EFL teachers prioritised linguistic accuracy in comparison to ESL teachers.

The analysis also indicated that short-term treatments and focused feedback seemed to produce better results. Both the total number of treatments as well as the number of treatments before the immediate posttest demonstrated that one-shot treatments yielded a noticeably larger effect than longer treatments. Upon further probing into the nine studies which had three or more treatment sessions, it was found that only one examined focused feedback, and two examined focused vs. unfocused feedback; the other six studies (67%) were unfocused. It could be that the wider scope adopted by these longer-term studies is often less salient, thus producing a smaller effect size. Conversely, short-term treatments tend to be more focused with more noticeable effects. Nonetheless, it is important to note that the duration of feedback may have had a bearing on other design features such as the feedback type, its intensity, and complexity of linguistic structure (Li, 2010).

A cross-tabulation of the findings of publication type with other independent variables revealed that only one out of the five dissertation studies (20%) had a one-shot treatment. As noted earlier, studies with shorter treatments tend to yield larger effect sizes. This is likely to be part of the reason why unpublished studies had a much smaller effect size than published ones.

In terms of the timing of publication, recent studies in the last 12 years (from 2008 onwards) produced a significantly larger effect than older ones. Explanations were sought and it was found that all the studies before 2008 are in the second language setting, whereas more than half of the studies from 2008 are in the foreign language context. This corresponds to the earlier findings that studies in the foreign language setting exhibited a greater effect size than studies in the second language setting. Moreover, with increased knowledge in research and reporting practices, recent studies have better-informed designs, thus possibly resulting in a larger effect size.

While the above discussion presents an overview of the findings, it is necessary to recall that the primary studies selected for this meta-analysis are heterogeneous, and the categories under the moderator variables may also not be equally represented. For instance, under the category of instructional status, there was only one study each in the primary school and junior college setting, and three in the secondary school setting, whereas the language programme and university settings featured more than 10 studies each. This uneven data set may not be an accurate representation of the specific context. Thus, results should be interpreted judiciously.

## **Conclusion**

By reviewing existing research and investigating the effect sizes across primary studies, this meta-analysis has attempted to consolidate findings of previous studies and provide more clarity on the effects of written corrective feedback on L2 writing. It was found that the overall

effective size is moderate, suggesting a positive influence of written corrective feedback. While differences between the effect sizes of direct or indirect feedback were not statistically significant, learners' proficiency stood out as the strongest mitigator. Thus, the message is clear – written corrective feedback can bring about improvement in L2 written accuracy. Teachers should continue to provide such feedback, and more importantly, take into account learners' developmental readiness to fully maximise the effectiveness of the feedback. Additionally, both direct and indirect feedback can benefit learners, and these function in tandem with other factors such as the treatment length and learners' proficiency.

This analysis has also surfaced possibilities for future research. First, more studies are needed in primary and secondary settings, as well as for learners of languages than English. Second, there is potential to expand the scope of investigation to uncover more mitigating factors such as learner motivation, teacher competencies, genre of writing tasks and research setting (laboratory-based or classroom-based). These variables are also likely to impact the efficacy of feedback. Third, it might be worthwhile to examine results of subsequent delayed posttests in addition to the first. These could potentially be a reliable measure of long-term effects. After all, sustainability and long-term improvements in the accurate use of the language are the ultimate goals of language learning. Finally, if coupled with a systematic review to complement the statistical results, the overarching findings would be even more robust and meaningful.

### About the Authors

**See Chen Lim** is a research student at the National Institute of Education, Nanyang Technological University, Singapore. She holds a Master of Educational Management and a Master of Arts (Applied Linguistics). She is an experienced educator who has been involved in teaching and curriculum planning as well as policy development. Her area of influence ranges from preschool to tertiary level. Her research interests include learner engagement and language teaching.

**Willy A Renandya** is a language teacher educator with extensive teaching experience in Asia. He currently teaches applied linguistics courses at the National Institute of Education, Nanyang Technological University, Singapore. He has given more than 100 presentations at regional and international ELT conferences and has published extensively in the area of second language education. He serves as an executive board member of Extensive Reading Foundation.

### References

- \*Ahmadi, D., Maftoon, P., & Mehrdad, A. G. (2012). Investigating the effects of two types of feedback on EFL students' writing. *Procedia – Social and Behavioral Sciences*, 46, 2590–2595. <https://doi.org/10.1016/j.sbspro.2012.05.529>
- Al-Jarrah, R. S. (2016). A suggested model of corrective feedback provision. *Ampersand*, 3, 98–107. <https://doi.org/10.1016/j.amper.2016.06.003>
- Biber, D., Nekrasova, T., & Horn, B. (2011). The effectiveness of feedback for L1-English and L2-writing development: a meta-analysis. *ETS Research Report Series*, 2011(1), i–99. <https://doi.org/10.1002/j.2333-8504.2011.tb02241.x>
- Bitchenor, J., & Knoch, U. (2008). The value of a focused approach to written corrective feedback. *ELT Journal*, 63(3), 204–211. <https://doi.org/10.1093/elt/ccn043>

- \*Bitchener, J., & Knoch, U. (2009). The contribution of written corrective feedback to language development: a ten month investigation. *Applied Linguistics*, 31(2), 193–214. <https://doi.org/10.1093/applin/amp016>
- \*Bitchener, J. (2008). Evidence in support of written corrective feedback. *Journal of Second Language Writing*, 17(2), 102–118. <https://doi.org/10.1016/j.jslw.2007.11.004>
- \*Bitchener, J., & Knoch, U. (2008). The value of written corrective feedback for migrant and international students. *Language Teaching Research*, 12(3), 409–431. <https://doi.org/10.1177/1362168808089924>
- Bitchener, J., & Knoch, U. (2009). The relative effectiveness of different types of direct written corrective feedback. *System*, 37(2), 322–329. <https://doi.org/10.1016/j.system.2008.12.006>
- \*Bitchener, J., & Knoch, U. (2010). Raising the linguistic accuracy level of advanced L2 writers with written corrective feedback. *Journal of Second Language Writing*, 19(4), 207–217. <https://doi.org/10.1016/j.jslw.2010.10.002>
- Bitchener, J., & Knoch, U. (2015). Written corrective feedback studies: approximate replication of Bitchener & Knoch (2010a) and Van Beuningen, De Jong & Kuiken (2012). *Language Teaching*, 48(3), 405–414. <https://doi.org/10.1017/s0261444815000130>
- \*Bitchener, J., Young, S., & Cameron, D. (2005). The effect of different types of corrective feedback on ESL student writing. *Journal of Second Language Writing*, 14(3), 191–205. <https://doi.org/10.1016/j.jslw.2005.08.001>
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2005). Comprehensive meta-analysis (version 3). Biostat.
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2011). *Introduction to meta-analysis* (2nd ed.; M. Borenstein, Ed.). Wiley.
- Carter, R., & Nunan, D. (Eds.). (2001). *The Cambridge Guide to Teaching English to Speakers of Other Languages* (The Cambridge Guides). Cambridge University Press. doi:10.1017/CBO9780511667206
- \*Chandler, J. (2003). The efficacy of various kinds of error feedback for improvement in the accuracy and fluency of L2 student writing. *Journal of Second Language Writing*, 12(3), 267–296. [https://doi.org/10.1016/s1060-3743\(03\)00038-9](https://doi.org/10.1016/s1060-3743(03)00038-9)
- Chen, J., Lin, J., & Jiang, L. (2016). Corrective feedback in SLA: theoretical relevance and empirical research. *English Language Teaching*, 9(11), 85–94. <https://doi.org/10.5539/elt.v9n11p85>
- Cooper, H., & Dent, A. (2011). *Ethical issues in the conduct and reporting of meta-analysis*. In A. T. Panter & S. K. Sterba (Eds.), *Multivariate applications series. Handbook of ethics in quantitative methodology* (pp. 417–443). Routledge/Taylor & Francis Group.
- Crivos, M. B., & Luchini, P. L. (2012). A pedagogical proposal for teaching grammar using consciousness-raising tasks. *MJAL*, 4(3), 141–153.

- DeKeyser, R. (2003). Explicit and Implicit Learning. In C. Doughty, & M. H. Long (Eds.), *The Handbook of Second Language Acquisition* (pp. 313-348). Blackwell.  
<http://dx.doi.org/10.1002/9780470756492.ch11>
- DeKeyser, R. M. (2007a). Introduction: Situating the concept of practice. In R. M. DeKeyser (Ed.), *Practice in a second language: Perspectives from applied linguistics and cognitive psychology* (pp. 1-18). Cambridge University Press.
- DeKeyser, R. (2007b). Skill acquisition theory. In B. VanPatten & J. Williams (Eds.), *Theories in second language acquisition: An introduction* (pp. 97-113). Lawrence Erlbaum Associates, Inc.
- Ellis, R. (2005). Principles of instructed language learning. *System*, 33(2), 209–224.  
<https://doi.org/10.1016/j.system.2004.12.006>
- Ellis, R. (2008). A typology of written corrective feedback types. *ELT Journal*, 63(2), 97–107. <https://doi.org/10.1093/elt/ccn023>
- Ellis, R. (2010). Epilogue. *Studies in Second Language Acquisition*, 32(2), 335–349.  
<https://doi.org/10.1017/s0272263109990544>
- Ellis, R. (2018). Meta-analysis in second language acquisition research. *Journal of Second Language Studies*, 1(2), 231–253. <https://doi.org/10.1075/jsls.00002.ell>
- \*Ellis, R., Sheen, Y., Murakami, M., & Takashima, H. (2008). The effects of focused and unfocused written corrective feedback in an English as a foreign language context. *System*, 36(3), 353–371. <https://doi.org/10.1016/j.system.2008.02.001>
- Evans, N. W., Hartshorn, K. J., McCollum, R. M., & Wolfersberger, M. (2010). Contextualizing corrective feedback in second language writing pedagogy. *Language Teaching Research*, 14(4), 445–463. <https://doi.org/10.1177/1362168810375367>
- \*Evans, N. W., James Hartshorn, K., & Strong-Krause, D. (2011). The efficacy of dynamic written corrective feedback for university-matriculated ESL learners. *System*, 39(2), 229–239.  
<https://doi.org/10.1016/j.system.2011.04.012>
- \*Farrokhi, F., & Sattarpour, S. (2012). The effects of direct written corrective feedback on improvement of grammatical accuracy of high-proficient L2 learners. *World Journal of Education*, 2(2). <https://doi.org/10.5430/wje.v2n2p49>
- \*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. [https://doi.org/10.1016/s1060-3743\(01\)00042-x](https://doi.org/10.1016/s1060-3743(01)00042-x)
- Ferris, D. R. (2006). Does error feedback help student writers? New evidence on the short- and long-term effects of written error correction. In K. Hyland & F. Hyland (Eds.), *Feedback in Second Language Writing: Contexts and Issues* (pp. 81–104). Cambridge University Press.
- Ferris, D. R. (2012). Written corrective feedback in second language acquisition and writing studies. *Language Teaching*, 45(04), 446–459. <https://doi.org/10.1017/s0261444812000250>
- Ferris, D. R. (2015). Written corrective feedback in L2 writing: Connors & Lunsford (1988); Lunsford & Lunsford (2008); Lalande (1982). *Language Teaching*, 48(04), 531–544.  
<https://doi.org/10.1017/s0261444815000257>

- Ferris, D. R., Liu, H., Sinha, A., & Senna, M. (2013). Written corrective feedback for individual L2 writers. *Journal of Second Language Writing*, 22(3), 307–329. <https://doi.org/10.1016/j.jslw.2012.09.009>
- \*Ferris, D. R., & Roberts, B. (2001). Error feedback in L2 writing classes. *Journal of Second Language Writing*, 10(3), 161–184. [https://doi.org/10.1016/s1060-3743\(01\)00039-x](https://doi.org/10.1016/s1060-3743(01)00039-x)
- \*Frear, D., & Chiu, Y. (2015). The effect of focused and unfocused indirect written corrective feedback on EFL learners' accuracy in new pieces of writing. *System*, 53, 24–34. <https://doi.org/10.1016/j.system.2015.06.006>
- \*Guo, Q., & Barrot, J. S. (2019). Effects of metalinguistic explanation and direct correction on EFL learners' linguistic accuracy. *Reading & Writing Quarterly*, 35(3), 261–276. <https://doi.org/10.1080/10573569.2018.1540320>
- Han, Y., & Hyland, F. (2015). Exploring learner engagement with written corrective feedback in a Chinese tertiary EFL classroom. *Journal of Second Language Writing*, 30, 31–44. <https://doi.org/10.1016/j.jslw.2015.08.002>
- \*Hartshorn, K. J., & Evans, N. W. (2015). The effects of dynamic written corrective feedback: a 30-week study. *Journal of Response to Writing*, 1(2). Retrieved from <https://journalrw.org/index.php/jrw/article/view/45>
- \*Hartshorn, K. J., Evans, N. W., Merrill, P. F., Sudweeks, R. R., Strong-Krause, D., & Anderson, N. J. (2010). Effects of dynamic corrective feedback on ESL writing accuracy. *TESOL Quarterly*, 44(1), 84–109. <https://doi.org/10.5054/tq.2010.213781>
- \*Hosseini, S. B. (2015). Written corrective feedback and the correct use of definite/indefinite articles. *International Journal on New Trends in Education and Their Implications*, 6(4), 98–112.
- Hyland, K., & Hyland, F. (2006). Feedback on second language students' writing. *Language Teaching*, 39(2), 83–101. <https://doi.org/10.1017/s0261444806003399>
- Hyland, K., & Hyland, F. (2012). *Feedback in second language writing : contexts and issues*. Cambridge: Cambridge University Press.
- Jafarigohar, M., & Jalali, M. (2014). The effects of processing instruction, consciousness-raising tasks, and textual input enhancement on intake and acquisition of the English causative structures. *Iranian Journal of Applied Linguistics*, 17(1), 93–118.
- \*Jhowry, K. (2010). *Does the provision of an intensive and highly focused indirect corrective feedback lead to accuracy?* (Unpublished Master's Dissertation). University of North Texas, ProQuest Dissertations Publishing.
- \*Jiang, L., & Xiao, H. (2014). The efficacy of written corrective feedback and language analytic ability on Chinese learners' explicit and implicit knowledge of English articles. *English Language Teaching*, 7(10), 22–34. <https://doi.org/10.5539/elt.v7n10p22>
- Kang, E., & Han, Z. (2015). The efficacy of written corrective feedback in improving L2 written accuracy: a meta-analysis. *The Modern Language Journal*, 99(1), 1–18. <https://doi.org/10.1111/modl.12189>

- \*Karim, K. M. R. (2014). *The effects of direct and indirect written corrective feedback (CF) on English-as-a-second-language (ESL) students' revision accuracy and writing skills* (Unpublished doctoral dissertation). Retrieved from <http://hdl.handle.net/1828/5157>
- \*Kassim, A., & Ng, L. L. (2014). Investigating the efficacy of focused and unfocused corrective feedback on the accurate use of prepositions in written work. *English Language Teaching*, 7(2), 119–130. <https://doi.org/10.5539/elt.v7n2p119>
- Kepner, C. G. (1991). An experiment in the relationship of types of written feedback to the development of second-language writing skills. *The Modern Language Journal*, 75, 305–313. <http://dx.doi.org/10.2307/328724>
- \*Khanlarzadeh, M., & Nemati, M. (2016). The effect of written corrective feedback on grammatical accuracy of EFL students: An improvement over previous unfocused designs. *Iranian Journal of Language Teaching Research*, 4(2), 55–68.
- Lee, I. (2012). Research into practice: Written corrective feedback. *Language Teaching*, 46(01), 108–119. <https://doi.org/10.1017/s0261444812000390>
- Li, S. (2010). The effectiveness of corrective feedback in SLA: a meta-analysis. *Language Learning*, 60(2), 309–365. <https://doi.org/10.1111/j.1467-9922.2010.00561.x>
- Li, S., & Vuono, A. (2019). Twenty-five years of research on oral and written corrective feedback in System. *System*, 84, 93–109. <https://doi.org/10.1016/j.system.2019.05.006>
- Liu, Q., & Brown, D. (2015). Methodological synthesis of research on the effectiveness of corrective feedback in L2 writing. *Journal of Second Language Writing*, 30, 66–81. <https://doi.org/10.1016/j.jslw.2015.08.011>
- Mackey, A., & Goo, J. (2007). Interaction research in SLA: A meta-analysis and research synthesis. In *Conversational interaction in second language acquisition* (pp. 407–453). Oxford University Press.
- \*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.
- Nassaji, H. (2018). Corrective feedback. In *The TESOL Encyclopedia of English Language Teaching* (pp. 1–7). <https://doi.org/10.1002/9781118784235.eelt0050>
- \*Nemati, M., Alavi, S. M., & Mohebbi, H. (2019). Assessing the effect of focused direct and focused indirect written corrective feedback on explicit and implicit knowledge of language learners. *Language Testing in Asia*, 9(1). <https://doi.org/10.1186/s40468-019-0084-9>
- Norris, J. M., & Ortega, L. (2000). Effectiveness of L2 instruction: a research synthesis and quantitative meta-analysis. *Language Learning*, 50(3), 417–528. <https://doi.org/10.1111/0023-8333.00136>
- Ortega, L., & Norris, J. M. (2006). *Synthesizing research on language learning and teaching*. John Benjamins Publishing Company.
- Polio, C. (2012). The relevance of second language acquisition theory to the written error correction debate. *Journal of Second Language Writing*, 21(4), 375–389. <https://doi.org/10.1016/j.jslw.2012.09.004>

- \*Rezazadeh, M., Tavakoli, M., & Rasekh, A. E. (2015). The effects of direct corrective feedback and metalinguistic explanation on EFL Learners' implicit and explicit knowledge of English definite and indefinite articles. *Journal of English Language Teaching and Learning*, 7(16), 113–146.
- Robb, T., Ross, S., & Shortreed, I. (1986). Salience of feedback on error and its effect on EFL writing quality. *TESOL Quarterly*, 20, 83–96. <http://dx.doi.org/10.2307/3586390>
- Russell, J. V., & 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). John Benjamins.
- \*Samiei, M., Tam, S. S., & Rouhi, A. (2017). Efficacy of written corrective feedback in the short and long term. *Journal of Modern Languages*, 27, 107–127.
- \*Sarvestani, M. S., & Pishkar, K. (2015). The effect of written corrective feedback on writing accuracy of intermediate learners. *Theory and Practice in Language Studies*, 5(10), 2046–2052. <https://doi.org/10.17507/tpls.0510.10>
- Schmidt, R. (2010). Attention, awareness, and individual differences in language learning. In W. M. Chan, S. Chi, K. N. Cin, J. Istanto, M. Nagami, J. W. Sew, T. Suthiwan, & I. Walker (Eds.), *Proceedings of CLaSIC 2010, Singapore, December 2-4* (pp. 721–737). National University of Singapore, Centre for Language Studies.
- Semke, H. D. (1984). Effects of the red pen. *Foreign Language Annals*, 17, 195–202. <http://dx.doi.org/10.1111/j.1944-9720.1984.tb01727.x>
- \*Sheen, Y. (2007). The effect of focused written corrective feedback and language aptitude on ESL learners' acquisition of articles. *TESOL Quarterly*, 41(2), 255–283. <https://doi.org/10.1002/j.1545-7249.2007.tb00059.x>
- \*Sheen, Y. (2010). Differential effects of oral and written corrective feedback in the ESL classroom. *Studies in Second Language Acquisition*, 32(2), 203–234. <https://doi.org/10.1017/s0272263109990507>
- \*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(4), 556–569. <https://doi.org/10.1016/j.system.2009.09.002>
- Sheppard, K. (1992). Two feedback types: Do they make a difference? *RELC Journal*, 23, 103–110. <http://dx.doi.org/10.1177/003368829202300107>
- Shin, H. W. (2010). Another look at Norris and Ortega (2000). *Working Papers in TESOL and Applied Linguistics*, 10(1), 15–38.
- \*Shintani, N., & Ellis, R. (2013). The comparative effect of direct written corrective feedback and metalinguistic explanation on learners' explicit and implicit knowledge of the English indefinite article. *Journal of Second Language Writing*, 22(3), 286–306. <https://doi.org/10.1016/j.jslw.2013.03.011>
- \*Shintani, N., Ellis, R., & Suzuki, W. (2013). Effects of written feedback and revision on learners' accuracy in using two English grammatical structures. *Language Learning*, 64(1), 103–131. <https://doi.org/10.1111/lang.12029>

Sia, P. F. D., & Cheung, Y. L. (2017). Written corrective feedback in writing instruction: a qualitative synthesis of recent research. *Issues in Language Studies*, 6(1), 61–80. <https://doi.org/10.33736/ils.478.2017>

\*Stefanou, C., & Révész, A. (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. <https://doi.org/10.1111/modl.12212>

Storch, N. (2010). Critical feedback on written corrective feedback research. *International Journal of English Studies*, 10(2), 29–46. <https://doi.org/10.6018/ijes/2010/2/119181>

\*Sun, S. (2013). *Written corrective feedback: Effects of focused and unfocused grammar correction on the case acquisition in L2 German* (Unpublished doctoral dissertation). Retrieved from <http://hdl.handle.net/1808/12284>

Truscott, J. (2007). The effect of error correction on learners' ability to write accurately. *Journal of Second Language Writing*, 16(4), 255–272. <https://doi.org/10.1016/j.jslw.2007.06.003>

\*Truscott, J., & Hsu, A. Y. (2008). Error correction, revision, and learning. *Journal of Second Language Writing*, 17(4), 292–305. <https://doi.org/10.1016/j.jslw.2008.05.003>

Van Beuningen, C. G. (2010). Corrective feedback in L2 writing: theoretical perspectives, empirical insights, and future directions. *International Journal of English Studies*, 10(2), 1–27. <https://doi.org/10.6018/ijes/2010/2/119171>

\*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. <https://doi.org/10.1075/itl.156.24beu>

\*Van Beuningen, C. G., De Jong, N. H., & Kuiken, F. (2011). Evidence on the effectiveness of comprehensive error correction in second language writing. *Language Learning*, 62(1), 1–41. <https://doi.org/10.1111/j.1467-9922.2011.00674.x>

Vygotsky, L. S. (1978). *Mind in Society: the Development of Higher Psychological Processes*. Harvard University Press.

Vygotsky, L.S. (1981). The genesis of higher mental functions. In J.V. Wertsch (ed.) *The Concept of Activity in Soviet Psychology*, 144–188. Armonk, NY: M.E. Sharpe.

Wang, T., & Jiang, L. (2014). Studies on written corrective feedback: theoretical perspectives, empirical evidence, and future directions. *English Language Teaching*, 8(1), 110–120. <https://doi.org/10.5539/elt.v8n1p110>

\*Wang, X. (2017). The effects of corrective feedback on Chinese learners' writing accuracy: a quantitative analysis in an EFL context. *World Journal of Education*, 7(2), 74–88. <https://doi.org/10.5430/wje.v7n2p74>

\*38 studies included in the current meta-analysis

## Appendix A

### Meta-analyses related to the Effectiveness of Corrective Feedback

Study	Period of Coverage	Mode of Corrective Feedback	No. of Studies ( <i>k</i> )	Outcomes	Specific Effect Size	Overall Effect Size
Norris & Ortega (2000)	1980 – 1998	Oral	49	L2 instruction	0.81 (recast) 0.96 (metalinguistic feedback)	–
Russell & Spada (2006)	1988 -2003	Oral & Written	15	L2 grammar learning	0.91 (oral) 1.31 (written)	1.16
Mackey & Goo (2007)	1990 – 2006	Oral	26	L2 oral proficiency	0.71 (immediate posttest) 1.09 (short-term delayed posttest)	–
Truscott (2007)	1980 – 2006	Written	13	L2 written accuracy	0.148 (uncontrolled studies) – 0.155 (controlled studies)	–
Li (2010)	1988 – 2007	Mostly Oral	33	L2 grammar acquisition	–	0.61 (fixed effects) 0.64 (random effects)
Biber et al. (2011)	1982 – 2007	Written & Oral	23	L1 & L2 writing proficiency and development	<u>Pre/posttest design</u> 0.86 (oral); 0.68 (written) <u>Treatment/control design</u> 0.84 (oral); 0.40 (written)	–
Kang & Han (2015)	1980 – 2013	Written	24	L2 written accuracy	–	0.54

[Back to the text](#)

## Appendix B

### Moderator Variables of Included Studies

No	Primary Studies	Setting	Target Lang	Instructional Status	Proficiency	Scope	Type	Total no. of Treatments	Treatments before Immediate Posttest	Type of Publication	Year of Publication
1	Fazio (2001)	SL	Non-Eng	Primary	NA	Unfocused	Direct	>3	>3	Published articles	Before 2008
2	Ferris & Roberts (2001)	SL	Eng	University	High Intermediate	Unfocused	Indirect	1	1	Published articles	Before 2008
3	Chandler (2003)	SL	Eng	University	Intermediate	Unfocused	Direct	4	4	Published articles	Before 2008
4	Bitchener (2005)	SL	Eng	Language Programme	High Intermediate	Focused	Direct	3	1	Published articles	Before 2008
5	Sheen (2007)	SL	Eng	Language Programme	Intermediate	Focused	Direct	2	2	Published articles	Before 2008
6	Truscott & Hsu (2008)	FL	Eng	University	Intermediate	Unfocused	Indirect	1	1	Published articles	From 2008
7	Bitchener & Knoch (2008)	SL	Eng	Language Programme	Low Intermediate	Focused	Direct	1	1	Published articles	From 2008
8	Bitchener (2008)	SL	Eng	Language Programme	Low Intermediate	Focused	Direct	2	1	Published articles	From 2008
9	Ellis et al. (2008)	FL	Eng	University	Intermediate	Focused vs unfocused	Direct	3	3	Published articles	From 2008
10	Beuningen et al. (2008)	SL	Non-Eng	Secondary	NA	Unfocused	Direct & Indirect	1	1	Published articles	From 2008
11	Sheen et al. (2009)	SL	Eng	Language Programme	Intermediate	Focused vs unfocused	Direct	2	2	Published articles	From 2008
12	Bitchener & Knoch (2010a)	SL	Eng	University	Low Intermediate	Focused	Direct	1	1	Published articles	From 2008
13	Bitchener & Knoch (2010b)	SL	Eng	University	Advanced	Focused	Indirect	1	1	Published articles	From 2008
14	Hartshorn et al. (2010)	SL	Eng	Language Programme	Advanced	Unfocused	Indirect	>3	>3	Published articles	From 2008
15	Sheen (2010)	SL	Eng	Language Programme	Intermediate	Focused	Direct	1	1	Published articles	From 2008
16	Jhowry (2010)	SL	Eng	University	NA	Focused	Indirect	4	4	Dissertation	From 2008
17	Evans et al. (2011)	SL	Eng	University	Advanced	Unfocused	Indirect	>3	>3	Published articles	From 2008
18	Beuningen et al. (2012)	SL	Non-Eng	Secondary	NA	Unfocused	Direct & Indirect	1	1	Published articles	From 2008
19	Farrokhi & Sattarpour (2012)	FL	Eng	University	Advanced	Focused vs unfocused	Direct	5	5	Published articles	From 2008
20	Ahmadi et al. (2012)	FL	Eng	University	Advanced	Focused	Direct & Indirect	6	6	Published articles	From 2008
21	Shintani & Ellis (2013)	SL	Eng	Language Programme	Low Intermediate	Focused	Direct	1	1	Published articles	From 2008
22	Murabak (2013)	FL	Eng	University	Low	Unfocused	Direct vs Indirect	12	12	Dissertation	From 2008
23	Karim (2013)	SL	Eng	Language Programme	Intermediate	Unfocused	Direct vs Indirect	3	1	Dissertation	From 2008
24	Kassim & Ng (2013)	FL	Eng	University	High Intermediate	Focused vs unfocused	Direct	1	1	Published articles	From 2008
25	Sun (2013)	FL	Non-Eng	University	NA	Focused vs Unfocused	Direct & Indirect	5	2 to 3	Dissertation	From 2008
26	Jiang & Xiao (2014)	FL	Eng	Junior College	Low Intermediate	Focused	Direct	2	2	Published articles	From 2008

27	Shintani, Ellis & Suzuki (2014)	FL	Eng	University	Low	Focused	Direct	1	1	Published articles	From 2008
28	Hartshorn & Evans (2015)	FL	Eng	Language Programme	Intermediate	Unfocused	Indirect	>3	>3	Published articles	From 2008
29	Stefanou & Révész (2015)	FL	Eng	Secondary School	Intermediate	Focused	Direct	2	2	Published articles	From 2008
30	Frear & Chiu (2015)	FL	Eng	University	Advanced	Focused vs Unfocused	Indirect	1	1	Published articles	From 2008
31	Rezazadeh (2015)	FL	Eng	University	Intermediate	Focused	Direct	1	1	Dissertation	From 2008
32	Hosseini (2015)	FL	Eng	Language Programme	Low	Focused	Direct vs Indirect	1	1	Published articles	From 2008
33	Khanlarzadeh & Nemati (2016)	FL	Eng	Language Programme	Low	Unfocused	Direct	3	1	Published articles	From 2008
34	Wang (2017)	FL	Eng	University	Advanced	Unfocused	Direct vs Indirect	1	1	Published articles	From 2008
35	Guo & Barrot (2019)	FL	Eng	University	Low	Focused	Direct	1	1	Published articles	From 2008

[Back to the text](#)

Copyright rests with authors. Please cite TESL-EJ appropriately.