An Analysis of Variation Sources in Written Corrective Feedback Studies: What is the Next Step?

Bonilla López, Marisela

An Analysis of Variation Sources in Written Corrective Feedback Studies: What is the Next Step?

Un análisis de las fuentes de variación en los estudios de retroalimentación correctiva escrita: ¿Cuál es el siguiente paso?

Marisela Bonilla López
Universidad de Costa Rica, Costa Rica
marisela.bonilla@ucr.ac.cr

http://orcid.org/0000-0002-1194-7721

Abstract:

While there is a significant amount of compelling evidence to support written corrective feedback (error correction), it is also acknowledged that research findings may not be applicable or conclusive enough given the great variability among studies. Nevertheless, no systematic attempt has been made to review and synthesize the extensive amount of literature to identify the sources that lead to such variation. This study aims to identify such variables. Results indicate that in a research base of 76 relevant publications, variations can be explained based on 11 main sources. Pursuant to these findings, this study sketches the main variance-related problems and outlines design recommendations to further expand L2 research and practice of error correction in Second Language Acquisition.

Keywords: Written corrective feedback, Sources of variation, L2 research and practice.

Resumen:

Existe una cantidad significativa de evidencia convincente a favor de la retroalimentación correctiva escrita (corrección de errores). Sin embargo, se reconoce que los resultados que arrojan las investigaciones quizás no sean tan pertinentes o lo suficientemente concluyentes debido a la gran variabilidad que hay entre los estudios. Aun así, no se ha hecho ningún intento sistemático de revisar y sintetizar la literatura de manera tal que se identifiquen las fuentes que causan tanta variación. Este artículo busca identificar esas variables. Los resultados indican que en una base de investigación que comprende 76 publicaciones relevantes, la variación puede explicarse a la luz de 11 fuentes principales. Este estudio también resume los problemas más significativos y propone recomendaciones de diseño para expandir la investigación y práctica de la corrección de errores en la enseñanza de una segunda lengua.

Palabras clave: Retroalimentación de corrección lingüística, Fuentes de variación, L2 Investigación y práctica.

1. Introduction

For many years, the benefits of correcting writing errors have been heavily contested (eg., Semke, 1984; Truscott, 1996, 2001). Such criticism prompted Second Language (L2) researchers to conduct further research on written corrective feedback (CF) (also known as error correction) and L2 composition teachers to continue to embrace a still ubiquitous practice. Studies on written CF have evolved significantly after more than forty years of research. Today, the effectiveness of written CF is no longer a cause for debate. In fact, research in L2 writing has demonstrated its potential as a revision tool (e.g., Ferris & Roberts, 2001) and L2 Acquisition studies have proven that written CF, actually, benefits L2 learning/development (e.g., Van Beuningen, De Jong, & Kuiken, 2012).

Despite notable progress made in research findings, different researchers (e.g., Guénette, 2007; Ferris, 2010) acknowledge that ample variability among studies impedes the attainment of firm conclusions. According to Russell and Spada (2006), "the constellation of moderating variables [makes it very difficult]

to establish clear patterns across studies" (p. 156). Similarly, Ellis, Sheen, Murakami, and Takashima (2008) claim that "research cannot provide language classroom teachers with clear-cut answers regarding what kind of CF to provide or how it should be provided. There are simply too many variables involved" (p. 567). Given this background, there is a need to identify and synthesize the root cause of such extensive variation of written CF; a review of this type could certainly contribute to determine exactly where such variance lies in the literature and what could be done to guarantee not only more sound comparisons across studies but also firmer conclusions.

Nonetheless, such a systematic analysis of current literature has not yet been carried out. Instead, existing meta-analyses (e.g., Kang & Han, 2015) and literature reviews (e.g., Truscott, 2007) have focused their attention either on analyzing the extent to which written CF studies address questions about L2 development (e.g., Bitchener, 2016), identifying the factors that may mediate the efficacy of feedback (e.g., Kang & Han, 2015) or on evaluating methodological and reporting practices (e.g., Liu & Brown, 2015). Therefore, the aim of this study is to identify variables that contribute to the extensive variability in error correction research.

For this purpose, this study attempts to answer the following question: What are the sources of variation in written CF studies? This analysis is based on a sample of 76 pre-selected publications (cf. Section 2) that describes factors that explain variability among studies (cf. Section 3) and outline variance-related problems followed by their respective design recommendations (cf. Section 4).

2. Methods

In a systematic literature review, "all available evidence on any given topic is retrieved and reviewed so that an overall picture of what is known about the topic is achieved" (Aveyard, 2010, p. 2), the task was guided by Preferred Items for Systematic Reviews and MetaAnalysis (PRISMA) statement (cf. Moher, Liberati, Tetzlaff, Altman, & The PrismaGroup, 2009).

2.1 Search Strategy

All studies on error correction were retrieved with a combination of *error, correction, written, writing, feedback, grammatical, grammar,* and *accuracy*. An online database search was also conducted (e.g., ProQuest, ERIC, JSTOR, Google Scholar, Scopus) as well as a journal search (e.g., *Journal of Second Language Writing, System, Language Learning, International Journal of Applied Linguistics*). A manual search of referenced relevant articles in previous meta-analyses was also conducted (e.g., Kang & Han, 2015).

2.2 Eligibility Criteria and Study Selection

The online and manual search resulted in 146 studies. Studies considered for this analysis were based on empirical efforts that targeted grammatical accuracy alone or in combination with other surface or textual-level issues to assess the short and long-term impact of feedback. For critical discussion purposes, it is important to note that the inclusion criteria also covered studies that did not include reports or a control group.

The analysis did not include: (1) unpublished MA theses or doctoral dissertations, (2) descriptive studies on learner views about error corrections, (3) studies investigating only oral CF, (4) studies that addressed written errors of learners exclusively with oral CF strategies, comments, or praise and criticism, (5) studies analyzing teacher response or learner response correctness, (6) studies regarding the differential effect of peer and teacher feedback, (7) studies in which feedback was provided strictly by peers or (8) studies endorsing

grading methods. After appraising the articles based on the inclusion/exclusion criteria, a total of 76 articles were selected to answer the research question that guided this systematic analysis.

3. RESULTS

Results indicate that the selected publications on written CF show 11 main sources of variation. This means that these variables are found in any given study, however with differing characteristics, their likelihood to impact accuracy, and their degree of detailed reporting. This section will list the identified variation sources, describe what they consist of, and quantify to what extent they are present in the research base.

3.1. Feedback Purposes

Feedback purpose refers to what the action of giving feedback actually promotes. Although written CF has been researched extensively for decades, the rationale behind examining improved accuracy in written CF has varied over the years (Polio, 2012). A useful distinction exists between *feedback for accuracy* and *feedback for acquisition* (see Manchón, 2011). Based on this classification, written CF promotes: I) immediate accuracy (i.e., feedback for accuracy); 2) the development of accuracy or study of second language acquisition (SLA) processes (i.e., feedback for acquisition) and 3) both short and long-term accuracy (i.e., feedback for accuracy and acquisition). Furthermore, research on written CF has mostly focused on feedback for acquisition studies (cf. Figure 1).

3.1.1. Feedback for Accuracy

Many studies focus solely on improving accuracy and improvement shown on the final rewritten version after feedback was given on the original draft (e.g., Fathman & Whalley, 1990). Two major characteristics are helpful in identifying studies about feedback for accuracy. One is their research agenda, which refers to the learning-to-write dimension (cf. Manchón, 2011), meaning that feedback for accuracy studies are directed at L2 writing with the goal of analyzing how learners can improve their self-editing and revision skills (Ferris, 2010). The other is design. Studies on feedback for accuracy refer to metrics for improved accuracy of revised texts (Bitchener, 2009) after comparing two versions of the same text. Although these studies are worth highlighting, since they are L2 writing oriented, some were conducted in classroom-based settings at the expense of methodological rigor. Lack of a control group proved to be the main shortcoming (e.g., Ferris & Roberts, 2001), raising concerns about the need for research designs that are both methodologically acceptable and pedagogically feasible (cf. Bitchener & Ferris, 2012).
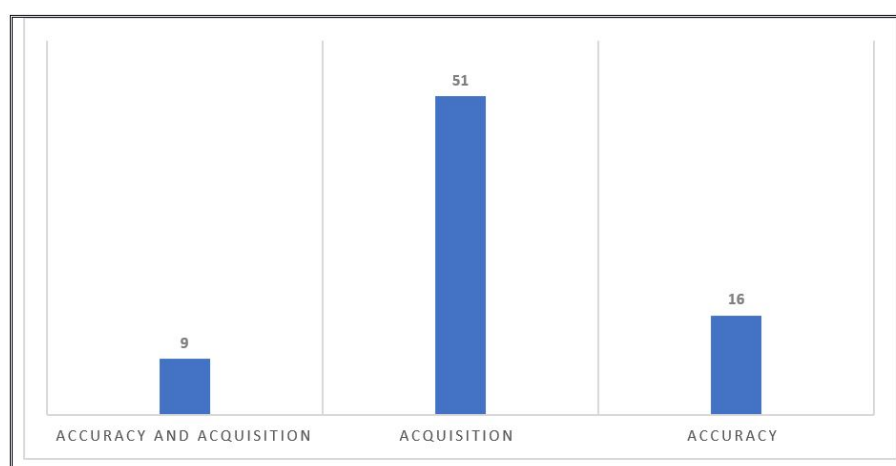
FIGURE 1
Number of Written CF Studies Based on Their Purpose
Source: Elaborated by author

3.1.2. Feedback for Acquisition

The origin of studies that measure feedback for acquisition can also be attributed to Truscott's (1996) controversial article which referred to grammar correction as "ineffective [and] harmful" (p. 327). After Truscott's strong opposition was further fueled in subsequent articles (Truscott, 2001, 2004, 2007), the hotly debated question now was: "Who is to say that short-term progress will be sustained over time?" (Ferris, 2004, p. 56). By questioning studies that measure feedback for accuracy given their lack of theoretical relevance (e.g., Truscott, 1996) or design and execution shortcomings (e.g., Guénette, 2007), a new strand of SLA-oriented feedback studies emerged.

Various features about feedback for acquisition studies can be drawn from research literature. Specifically, they explore the dimension of writing-to-learn-language (Manchón, 2011). This dimension is SLA-driven (Sheen, 2007), meaning that research aims at investigating "the manner in which writing (including both text generation activity and the processing of feedback) can aid interlanguage development" (Santos, López-Serrano, & Manchón, 2010, p. 132). Based on this research agenda, the objective of each feedback for acquisition study varies: (1) in testing the effect of written CF either on learner's ability to acquire language forms over time (Ferris, 2010) or (2) in exploring processes relevant for L2 acquisition (Van Beuningen, 2010). If interest is in L2 development (e.g., Bitchener & Knoch, 2010; Frear & Chiu, 2015; Shintani, Ellis, & Suzuki, 2014), the study design is likely to focus on the learner´s attention on a limited number of pre-selected categories, include pretest and posttest-delay measures, express interest in learning outcomes (e.g., L2 improvement over time) and discard revisions. Although still classified as feedback for acquisition, exceptions to the former characterization targeted a large array of errors (e.g, Ferris, 2006) or compared first and last writing pieces over time (e.g., Chandler, 2003).

Alternatively, when interest lies in SLA processes (e.g., Qi & Lapkin, 2001), study designs tend to target multiple errors, examine text revisions, prioritize the feedback process (Van Beuningen, 2010), and discard measures of improved accuracy in newly produced texts over a period of time. Overall, in 51 feedback for acquisition studies, only a few researched SLA processes (n = 9). The purpose of most feedback for acquisition studies has been to examine improvement of accuracy either by comparing first and final drafts throughout a semester (n = 22) or using a pretest-posttest-delayed design (n = 19).

3.1.3. Feedback for Accuracy and Acquisition

In 2010, Ferris described the ideal research design as one that is *blended*. A blended study design addresses "the L2 writing starting point (i.e., whether written CF helps students develop more effective revision and self-editing skills) as well as the more immediate SLA concern of whether written CF helps learners acquire a target structure" (Ferris, 2010, p. 195). One may argue, however, that studies with blended designs may be considered as studies where corrective feedback involves feedback for both accuracy and acquisition.

The main characteristic of feedback studies with blended designs is the comparison of the original text with its revised version to determine immediate accuracy and the use of longitudinal measures (i.e., pretest-posttest-delayed posttest) to assess long-term accuracy. Yet, 'early versions' of blended design studies did not include control groups or strictly adopt pretest-posttest-delayed posttest designs (e.g., Chandler, 2003; Ferris, 2006). Despite differences in their study designs, Truscott and Hsu (2008), van Beuningen et al. (2008, 2012), Diab (2015), Bonilla, Van Steendam, and Buyse (2017), as well as Bonilla, Van Steendam, Speelman, and Buyse (2018) conducted studies with more rigorous methodologies and controlled conditions.

3.2. Feedback Scope

Brown (2012) defines *feedback scope* as "the number and type of errors that are addressed-either with a comprehensive approach or a focus on a limited range of error categories" (p. 863). Specifically, different error types have been targeted in written CF studies, such as grammatical and non-grammatical errors (e.g., van Beuningen et al. 2012). Nevertheless, despite the type of error, written CF may include focused (i.e., selective) CF or unfocused (i.e., comprehensive) CF. Literature has generally referred to focused CF feedback as one that addresses "a single or a limited number" (Stefanou & Révész, 2015, p. 264) of linguistic categories, usually one or two. On the other hand, unfocused CF has been traditionally understood as feedback that corrects "all (or at least a range of) errors" (Ellis et al., 2008, p. 356). In a recent meta-analysis of written CF studies, Liu and Brown, (2015) proposed a classification that distinguished the scope by the exact number of targeted errors: highly focused (i.e., highly selective) (1 error type), mid-focused (i.e., mid selective) (2-5 error types), and unfocused CF (i.e., comprehensive) (> 5 error types). Table 1 shows the number of studies addressing any one scope type either in single form or by comparing them.

TABLE 1
Number of Written CF Studies per Scope

	n	Sample Publication
Highly Selective	19	Bitchener & Knoch (2010)
Mid-Selective	13	Diab (2015)
Comprehensive	40	Bonilla et al. (2017)
Highly-Selective vs Mid-Selective	2	Ellis et al. (2008)
Highly-selective vs. Comprehensive	2	Frear & Chiu (2015)
Total	76	

Elaborated by author

3.3. Feedback Type

Feedback type deals with the strategies used for providing written CF. The reviewed literature refers to eight different strategies which have been implemented as grouping feedback variables (i.e., in combination with others) or as single feedback variables (i.e., unsupplemented). Adapted from Ellis (2009), Table 2 lists the strategies available to correct learners' written errors and briefly describes them.

TABLE 2
Overview of Written CF Strategies

Strategy	Description
Direct Corrections	In-text provision of the correct form. The target form is inserted above the error and/or any erroneous language use is crossed out.
Reformulations	Provision of the correct forms on a separate individualized handout. Each learner's text is used to generate a native-like reformulated version.
Metalinguistic Codes	Labels (in abbreviated forms) which describe the nature of the error (e.g., WF for word form errors). They can be placed either above the error or in the margin.
Metalinguistic Rule Reminders	A (brief) metalinguistic explanation about the error. A number is usually placed above the error, which refers to the rule added at the bottom of the learner's text. They can also be provided in oral form (e.g., conferencing).
Underlining (circling)	Underlining (or circling) of the error. No further corrective information is provided in the text.
Marginal Check Marks	Marginal indication of an error. The feedback provider places on the margins of the text as many check marks ($\sqrt{}$) as there are errors in a line.
Concordances	Inserted hyperlinks next to the error. The link provides learners with concordance information about correct language usage.
Models	Provision of a separate generic text. A model writing piece for all learners is created by taking into consideration learner (e.g., proficiency level) and task (e.g., genre) variables.

Source: Adapted from Ellis (2009)

With respect to feedback strategies, the status of the field reveals that some have received more empirical attention than others. In Figure 2, direct corrections comprise the most examined strategy whereas concordances together with check marks and models have been largely overlooked.
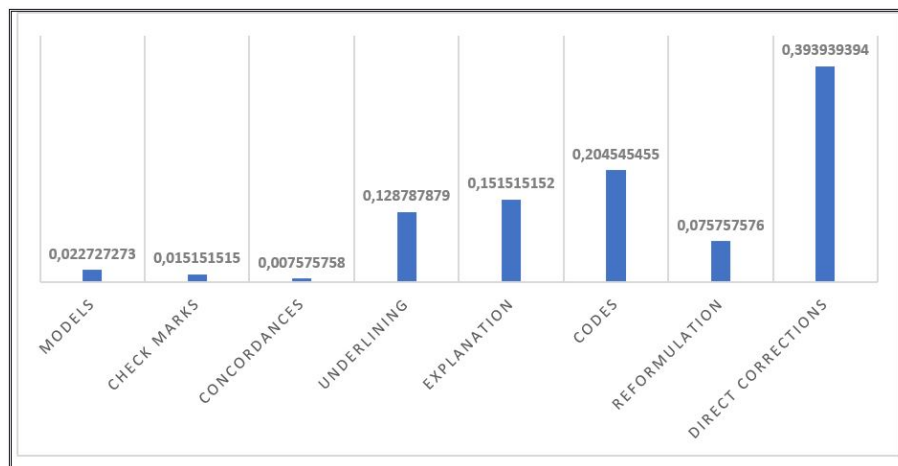
FIGURE 2
Percentage of Strategy Use in Written CF Studies
Source: Elaborated by author

3.4. Feedback Content

Feedback content refers to the issues that are the emphasis of the CF treatment. Written CF has been mostly conceived as feedback on linguistic issues. For this reason, studies on learner response to grammar correction are more prevalent in literature. Storch (2010) argues that "it is feedback on language use, termed written corrective feedback (WCF), which seems to have attracted the most research attention recently" (p. 30). Nevertheless, to a lesser degree the bulk of written CF studies has also included research that examines grammatical accuracy improvement by including groups that receive additional grammar instruction (e.g., Frantzen, 1995) or feedback on content (e.g., Fathman & Whalley, 1990; Fazio, 2001; Kepner, 1991; Sheppard, 1992). Other studies have treated grammar issues together with organization and content (e.g., Stiff, 1967) or lexical issues (e.g., Lalande, 1982). The content in feedback studies thus far can be seen in Figure 3.
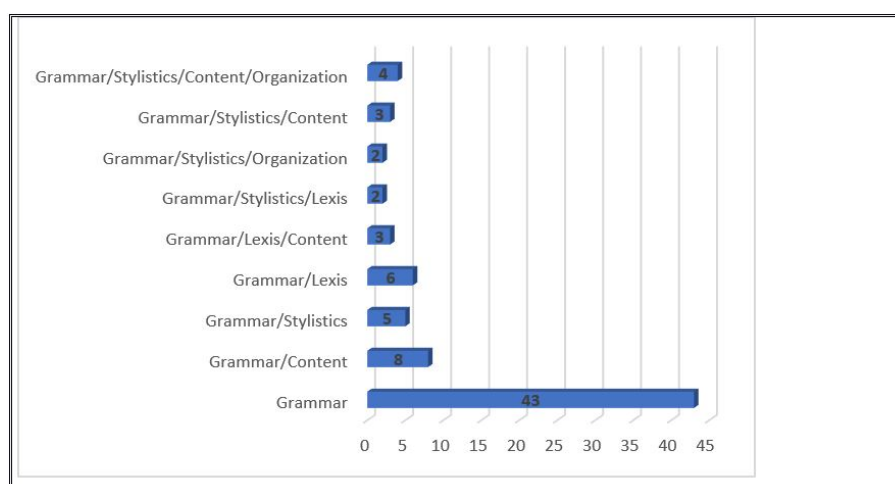

FIGURE 3.
Number of Written CF Studies by Content
Source: Elaborated by author

3.5. Feedback Frequency

Feedback frequency refers to the number of written CF sessions in a given study. Although there is no agreed-upon definition, Storch (2010) is the only researcher that makes a clear distinction regarding duration: a one-shot treatment vis-à-vis a sustained treatment. The author refers to one-shot design CF studies as those that give CF on one occasion only whereas sustained ones are those that give CF on more than one writing text.

   Based on this definition, sustained treatment studies are more common (see Figure 4), especially in early CF studies. They provided CF over the course of a term and (coincidentally) also targeted multiple errors (e.g., Chandler, 2003; Ferris, 2006). More recent sustained CF studies target either a few linguistic categories (e.g., Ferris, Liu, Sinha, Senna, 2013) or a broader range of categories (Lavolette, Polio, Kahng, 2015). Although the one-shot treatment was less prevalent in early studies (Fathman & Whalley, 1990; Ferris & Roberts, 2001), it has become the most popular design for SLA-oriented feedback studies and focuses either on a narrow number of preselected error categories (e.g., Stefanou & Révész, 2015) or on a large array of categories (Santos et al., 2010).
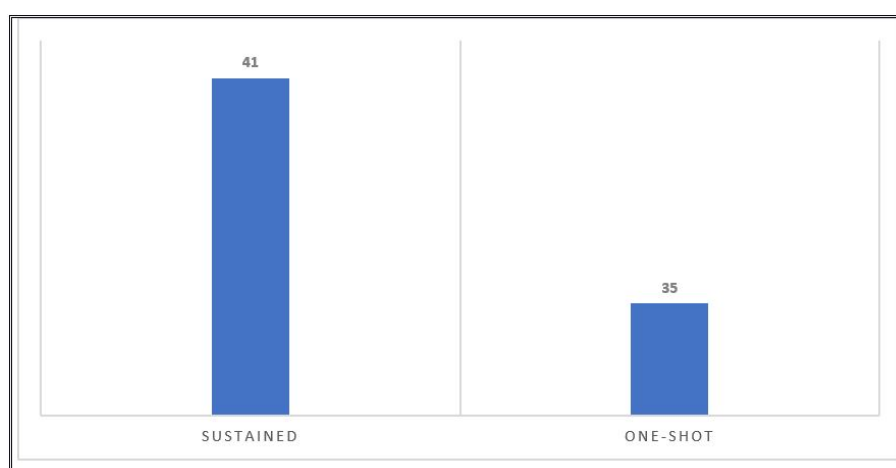

FIGURE 4.
Number of Written CF Studies by Frequency
Source: Elaborated by author

3.6. Feedback Elicitation Task

Another variable that causes wide disparity in CF studies is the type of writing assignment that is assessed which ranges from dicto-gloss tasks (e.g., Shintani et al., 2014) and journal entries (e.g., Fazio, 2001) to picture-based stories (Frear & Chiu, 2015) and fables (e.g., Farrokhi & Sattarpour, 2012). Figure 5 displays writing tasks in written CF research out of which three elicitation tasks stand out: picture-based narratives, picture-based description and argumentative texts.

   Upon interpreting the results, the broad gamut of elicitation tasks is a significant variable to consider. Murphy and De Larios (2010) argue that readers should remember that tasks can vary in terms of the restrictions imposed on learners and the levels of cognitive complexity of the tasks. For example, although both picture-based stories and essays are time-compressed tasks, they differ in "that the former [is] less cognitively complex in terms of conceptualization and linguistic encoding demands than the latter" (p. vii). This is also the case for picture description tasks which may vary with regards to the complexity of the pictures or the degree of expected precision in the description (Hanaoka & Izumi, 2012).
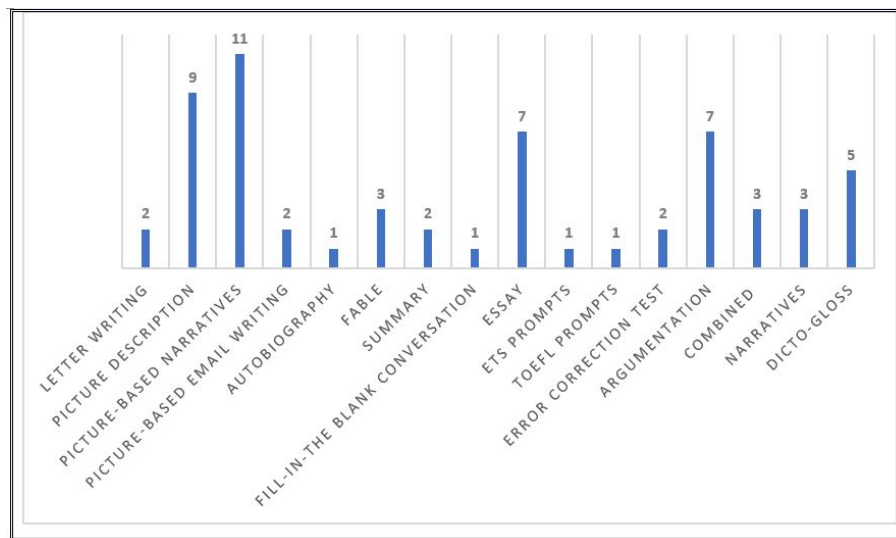
FIGURE 5
Number of Written CF Studies by Elicitation Task
Source: Elaborated by author

3.7. Feedback Source

The seventh source of variation pertains to the individual responsible for addressing the written errors of the student. Although studies on researcher-initiated feedback are predominant in research literature, the teacher is not the only source of written CF available to learners. Classmates, friends, tutors, researchers, the actual writers, and software programs (Ferris, 1997) may also be considered as additional sources of feedback. In our research, there are studies where the individual who provides feedback is a teacher (e.g., Abedi, Latifi, & Moinzadeh, 2010), a teacher-researcher (e.g., Ahmadi, Maftoon, & Mehrdad, 2012), or a software program (e.g., Lavolette et al., 2015), among others. Figure 6 summarizes current feedback sources in feedback literature.

The role of the corrector or the person responsible for correcting learner errors has been a matter of interest since early discussions (e.g., Hendrickson, 1978) and is still an underexplored and underreported variable. The feedback source matters since it could play a role in interpretation of the results. For example, Fazio (2001) considers that an affective factor concerning the feedback source plays a role in feedback outcome. As the author acknowledges, there might not have been a feedback effect because corrections were provided by a teacher other than the familiar one, "which may have had an effect on the manner in which students reacted to their feedback" (p. 245).
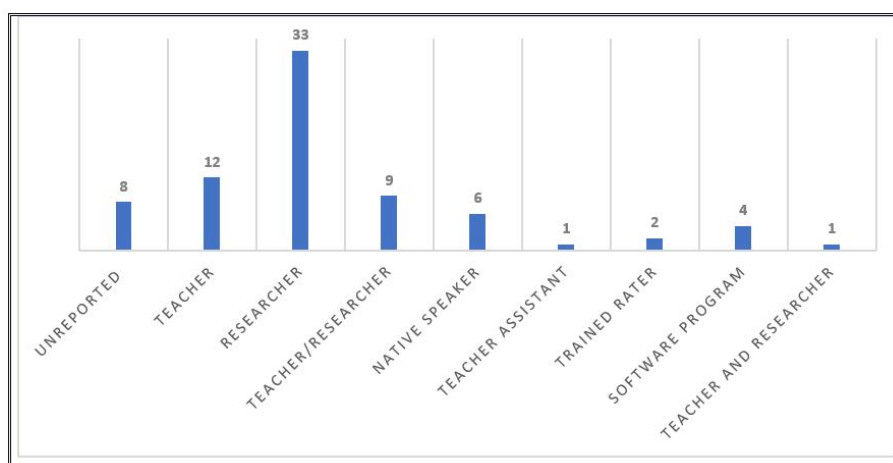
FIGURE 6
Number of Written CF Studies by Source
Source: Elaborated by author

3.8. Feedback Medium

Feedback medium refers to the different response modes to a learner's written output. Overall, the method for delivering written CF is pen-and-paper compositions, face-to-face interaction (e.g., conferencing) or computers. Most empirical evidence on the effects of written CF on learners' revised and/or new texts originate from studies that researched the differential effect of various CF techniques and were delivered in the same manner, namely on paper (e.g., Bonilla et al., 2017) (see Figure 7). Also, learners' written inaccuracies have been addressed with combined pen-and-paper compositions as well as conferencing (e.g., Bitchener & Knoch, 2009a, 2009b). Computers, which according to Hyland and Hyland (2006), allow synchronous (i.e., in real time) or asynchronous (i.e., delayed) communication, have also been used to observe learners' written grammatical accuracy, though to a lesser degree. Computers have been used to generate or mediate CF such as automated text evaluation (e.g., El-Ebyary & Windeatt, 2010; Nagata, 1997), for example, via concordance files (e.g., Gaskell & Cobb, 2004) or Google Docs (e.g., Shintani & Aubrey, 2016).
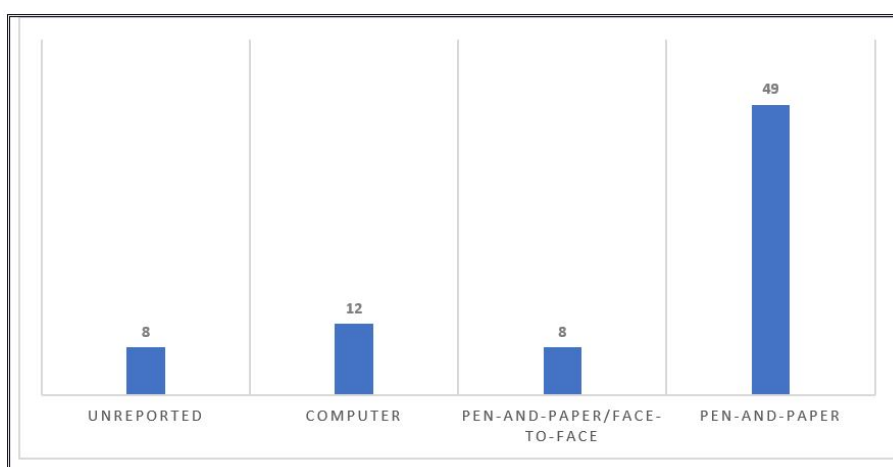

FIGURE 7
Number of Written CF Studies by Medium
Source: Elaborated by author

Number of Written CF Studies by Medium

Source: Elaborated by author

Advances in technology and, particularly, computer-assisted CF, have made it possible for research to be conducted in other feedback medium not available before (e.g., effect of feedback timing). Extensive research by authors (Lavolette et al. (2015), Shintani (2015), Shintani and Aubrey (2016)) now provides evidence of the effect of synchronous CF (i.e., feedback while learners compose a text) as well as asynchronous CF (i.e., feedback after text completion). Novel studies by the aforementioned authors compare differing feedback conditions provided via the same medium (i.e., via computer). Yet, less empirical data can be obtained, regarding the differential effect of various feedback modalities (but see Sheen, 2010).

3.9. Feedback Processing

Additional variation is observed from student response to written CF. Feedback processing refers to how students use the feedback that is provided. As Ellis (2009) explains, students can be asked to deal with written CF in three different ways: (1) respond to the feedback and revise the text, (2) address the feedback without making corrections or (3) refuse to address the feedback (i.e., receiving CF with no further action required). Figure 8 shows the extent that feedback processing form has been implemented in feedback literature.
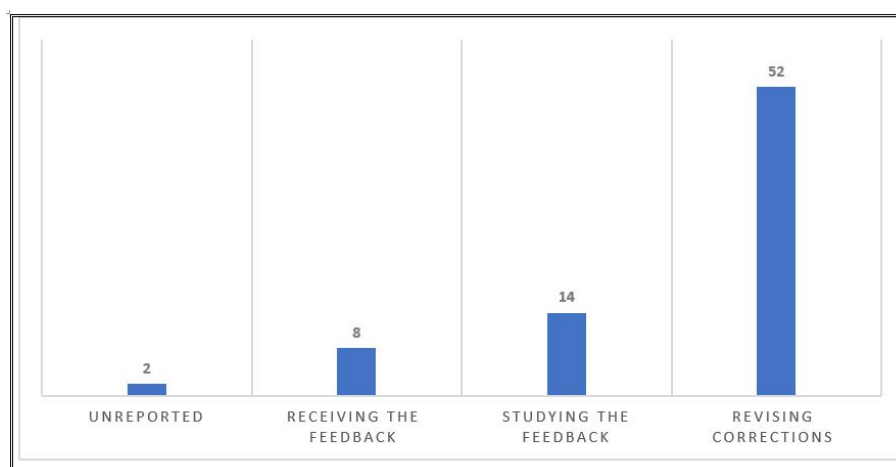
Figure 8.

FIGURE 8
Number of Studies per Feedback Processing
Source: Elaborated by author

3.9.1. Process the Feedback by Revising

In most research on written CF, students viewed the feedback and self-corrected their errors. In fact, text revision was the most prevalent self-correction activity. Researchers consider that revisions are a useful intervention that enhance the effect of feedback (e.g., Shintani et al., 2014) and may be a necessary step for L2 development (e.g., Ferris, 2004) since it benefits writers (e.g., Hedgcock & Lefkowitz, 1992), constitutes an important stage in the writing process (e.g., Ellis, 2009), affords opportunities to refine acquired L2 knowledge (e.g., Chaudron, 1984), and provides pushed output (e.g., Shintani & Ellis, 2013). Overall, emphasis has been made about the need to increase the possibility for learners to process CF, which can only be achieved if L2 teachers "provide activities which reinforce students' attention to it" (MacDonald, 1991, p. 3). Nevertheless, this literature review revealed vast discrepancies in addressing corrections and

carrying out revisions. For example, in some studies (n = 28) learners read their feedback and revised their text accordingly with the feedback at hand (e.g., Sampson, 2012; van Beuningen et al., 2008, 2012). In other studies on text revision procedures (n = 14), learners viewed the written CF, although students could not refer to it for subsequent text revision (e.g., Martínez & De Larios, 2010) and in some cases (n = 10), the revision conditions were not stated clearly (e.g., Fathman & Whalley, 1990).

3.9.2. Process the Feedback by Analysis

Unlike studies which include an additional component in their design so that learners may self-correct errors (e.g., revision), others spend more time allowing learners to analyze feedback (n = 14) (e.g., Ellis et al., 2008). Most SLA-oriented feedback studies on L2 development do not include a review component. Overall, research has shown that drawing learner attention to form, by asking them to study the feedback, has sufficed to improve accuracy for specific L2 features. According to Shintani and Ellis (2015), "it is also possible for feedback to be effective even if there is no opportunity to revise, provided that learners be required to process corrections" (p. 111), meaning that despite its effectiveness, revisions are not the only type of functional intervention.

3.9.3. Receiving Feedback

The value of the third type of student response (n = 8) is, undoubtedly, the least interesting (theoretically and pedagogically) since written CF feedback is meaningless if learners are not required to respond to it (Storch, 2010), and only works if learners are able to take note and process it (Ellis, 2009). As previously mentioned, unless learners are required to do something, chances are that they will either throw the composition in the garbage (Guénette, 2012) or simply make a mental note of the feedback (cf. Cohen, 1987). Either way, if the learner does not act on the feedback, it may be considered as a wasted effort by the teacher, equal to never having been given any feedback at all (Chandler, 2003). Not surprisingly, among the four basic requirements for conducting written CF studies, Evans, Hartshorn, and Strong-Krause (2011) state that teachers must ensure learners respond to the feedback, contrasting with Fazio (2001), who acknowledged that a simple verbal reminder to review feedback given for a journal entry before starting on the next one may not suffice to verify learner response to feedback—hence the lack of feedback effect in the study.

3.10. Learner Variables

Differences in learner variables provide more sources of variation in written CF studies. Evans, Hartshorn, McCollum and Wolfersberger (2010) define learner variables as those "that the learner brings to the learning experience" (p. 448). Learner factors include "age, language aptitude, memory, learning style, preferences, personality, motivation, language anxiety, learner beliefs" (Ellis, 2010, p. 339), "nationality, cultural identity, learning style, values, attitudes, beliefs, socioeconomic background" (Evan el at., 2010, p. 448), "personality … goals and expectations, motivations, past learning experiences, preferred learning styles and strategies, content knowledge and interest, time constraints, attitudes towards the teacher, the class, the content, the writing assignment" (Goldstein, 2004, p. 67), and "reactions to teacher feedback and their investment in the course" (Hyland & Hyland, 2006, p. 88). Unfortunately, few learner factors have been addressed empirically and tend to be underreported. An example is learner's (language) proficiency level, acknowledged as a crucial learner variable that could determine the extent of feedback effectiveness (e.g., Bitchener & Knoch, 2010; Ferris, 1999). Although it has been suggested that written CF should be provided based on learner proficiency level (e.g., Guénette, 2007), it is not always possible to identify a participant's entry level

proficiency (n = 34) and, if stated, (n = 42), it is not always reported correctly. The problem then lies in the L2 teachers' inability to find practical applications in their classes and L2 writing or SLA researchers to replicate studies as emphasized by Guénette (2007).

The aspect that has received the most empirical consideration in literature regarding written CF is learner's attitude towards written CF (e.g., Bonilla et al., 2017) (see Figure 9). Other aspects include educational level (e.g., van Beuningen et al., 2012), language analytic ability (e.g., Benson & DeKeyser, 2018), grammatical sensitivity and knowledge of metalanguage (e.g., Stefanou & Révész, 2015) where only the last three were examined as mediating variables.
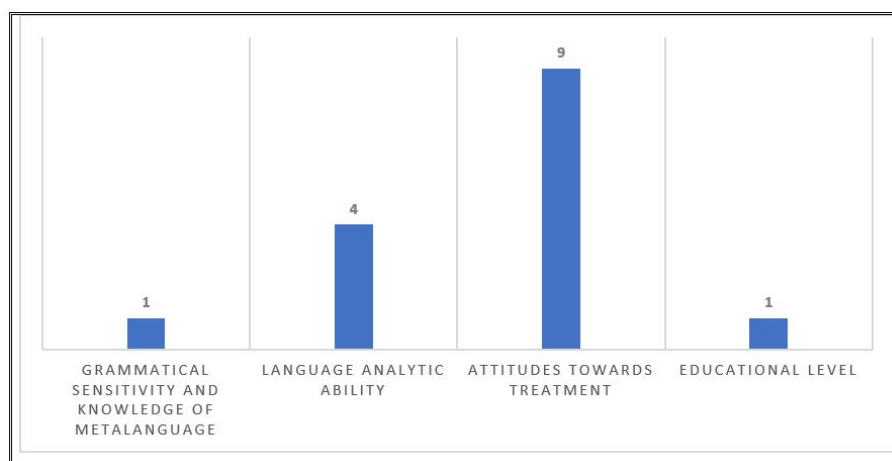
FIGURE 9
Number of written CF Studies Addressing Learner Factors
Source: Elaborated by author

3.11. Context

The last source of variation is the setting for instructional written CF. According to Ellis (2010), "contextual variables include … macro factors related to the setting in which learning takes place (e.g., immersion, foreign language [FL], and L2 settings)" (p. 341). The most common macro distinction for written CF studies are SL and FL contexts (Ellis, 2010; Murphy & de Larios, 2010). Figure 10 depicts various written CF settings. Most research includes studies conducted with ESL learners in the United States (USA) (e.g., Hartshorn et al., 2010) or in other English-dominant settings such as New Zealand (e.g., Bitchener, Young, & Cameron, 2005), Australia (e.g., Storch & Wigglesworth, 2010), Hong Kong (e.g., Lee, 1997), or Canada (e.g., Qi & Lapkin, 2001). There are also FL studies in the United States conducted with students learning Spanish (e.g., Frantzen, 1995; Kepner, 1991), German (e.g., Semke, 1984) and Japanese (e.g., Nagata, 1997).
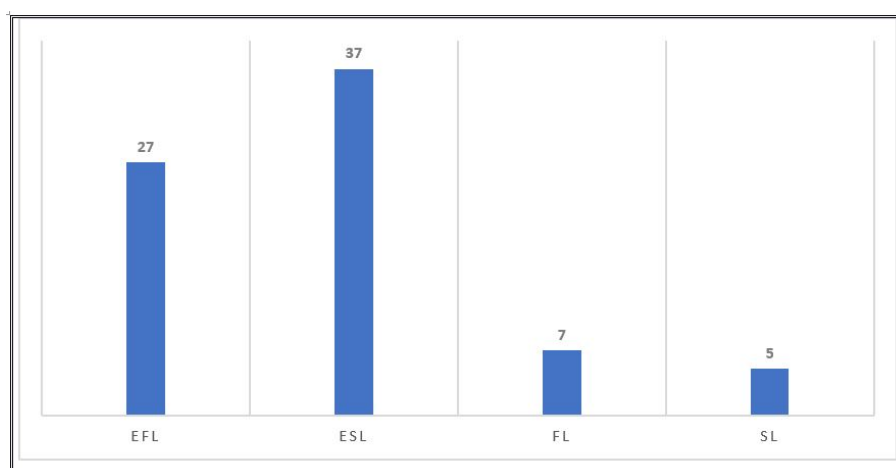
FIGURE 10.
Number of Written CF Studies by Macro-Context
Source: Elaborated by author

4. Discussion and Further Research

While identifying and quantifying the main sources of variability in the written CF research base may allow L2 researchers and practitioners alike to understand why some theoretical and/or practical questions remain partially unanswered, a brief discussion on variance-related issues is in order. Thus, the following section sketches the most significant problems gleaned from the systematic analysis of the literature (cf. Section 3), followed by their respective design recommendation.

4.1. Include a Revision Component.

Problem: A few feedback for acquisition studies require revision.

This may be understandable when the ultimate goal is the acquisition of L2 forms (cf. Ferris, 2004). Nevertheless, two reasons can be discerned regarding the importance of including revision in a design: they are the theoretical and pedagogical relevance for L2 writing practitioners. According to Ferris (2010), L2 composition teachers have not lost interest in determining strategies that help learners effectively edit and revise their work. Revision may also be useful for learners to acquire an L2 (Ferris, 2004) and may provide written CF researchers with more insight to further substantiate results. Therefore, literature may benefit from studies with blended designs to fill the void between theory and practice and yield results that could be pedagogically and theoretically relevant for two lines of inquiry: L2 Writing and SLA.

4.2. Conduct Studies in Underrepresented Settings.

Problem: Until now, a large bulk of written CF studies have taken place in English-dominant environments.

This is particularly problematic because "feedback is an area of work that affects all writing teachers and their students, [and as such] it is important that literature be augmented by research studies conducted in different parts of the world" (Lee, 2014, p. 2). It is paramount to stop assuming that FL learners outside and in the United States as well as SL learners in the United States and other English-dominant contexts are comparable (cf. Ferris, 2010). Clearly, L2 knowledge, motivation, and learning experiences (among others) that these types of learners bring to L2 writing may be quite different, potentially influencing how they engage

and benefit from feedback. Not surprisingly, Evans et al. (2011) add that the effectiveness of feedback may vary to such an extent that what works well in one context may not necessarily work well in another one. That is why further studies in underrepresented geographical areas are necessary to move this field forward and gain knowledge about feedback practices for particular settings.

4.3. Assign Ecologically Valid Writing Tasks.

Problem: A considerable number of studies do not consider the communicative purpose of an elicitation task (as a potential influential factor of feedback outcome) as well as its ecological validity (as a factor to determine the applicability of research findings).

An example related to the communicative purpose of the elicitation task is journal writing, which (contrary to its very intention) is thought to be discouraging if corrected. As Guénette (2007) explains, its purpose is "to encourage students to write and foster a comfortable and positive writing environment" (p. 47). Hence, targeting learners' written errors with such tasks may be counterproductive. Second, findings that emerge from a given study may also be analyzed from the perspective of their ecological validity and the extent to which the task itself is applicable. If tasks are not part of the curriculum, since they are meant for research purposes only (e.g., Bitchener & Knoch, 2010; Sheen, 2007; van Beuningen et al., 2012), the ecological validity of the elicitation task could be listed as another variable that hinders transferring findings from a laboratory-like setting to the actual L2 writing class context. Consequently, it is recommended to assign tasks that reflect those usually assigned in an L2 class and that allow freedom of expression (cf. Bruton, 2009a).

4.4. Report Individual and Contextual Variables Rigorously

Problem: Key design variables are vaguely reported or not mentioned at all.

A basic requirement for seeking useful answers to other instructional contexts is rigorous reporting of all variables. Nevertheless, research does not always reveal who provides feedback, who the learners are, how they write and receive the feedback, what they will do with it, and what they gain from engaging with it. These information gaps prevent firm conclusions to be drawn about the effectiveness of written CF. Along these lines, Guénette (2007) explains that,

> rather than interpret the conflicting results as a demonstration of the effectiveness or ineffectiveness of corrective feedback on form, I suggest that findings be attributed to the research design and methodology, as well as the presence of external variables beyond the control and vigilance of the researchers. (p. 40)

The importance of reporting all relevant individual and contextual factors in detail must, therefore, be addressed to obtain realistic and accurate conclusions.

4.5. Consider Learner Variables.

Problem: A large bulk of the research base focuses only on accuracy outcomes.

Responding to L2 learners' written errors is far more complex than drawing conclusions exclusively from product data. Ferris et al. (2013) argue that "[viewing] student texts in isolation do not provide researchers or teachers enough information about if/how/why WCF helps student writers improve (or does not). Rather, there must be careful consideration of the larger, classroom characteristics, the teacher and the learners themselves" (p. 324), which implies that further research is required to gain more insight into what learners bring to the feedback experience (cf. Evans et al., 2010) and how that could impact feedback outcome.

Therefore, empirical feedback studies that add a qualitative component to its design and address learner factors that have been neglected in empirical feedback studies such as learner proficiency level (Bitchener, 2012) would be a valuable addition to the research base.

4.6. Ensure that the main requirements for measuring accuracy are met

Problem: The measures for accuracy and design characteristics bear large discrepancies.

Part of the research on written CF includes studies that lack a control group (e.g., Asassfeh, 2013), used differing genres for pretest and (immediate and delayed) posttest comparisons (e.g, Diab, 2015), omitted tracing errors (e.g., van Beuningen et al., 2008) and had learners receive feedback without requiring that it be processed it in any way (e.g., Fazio, 2001). Nevertheless, researchers have described what the ideal design must have if the effectiveness of written CF is to be measured. Bitchener (2008) explains that posttest measurements are only valid if tasks are comparable with the pretest; the author then warns that varied genres must not be used. Other requirements include a control group and delayed posttest to be able to measure retention. Along similar lines, Bruton (2009b) adds that learners must be able to use the feedback received with targeted features tracked in subsequent writing, Thus, future studies have important issues to resolve in their attempt to measure accuracy without methodological criticism.

5. Conclusion and Limitations

The broad gamut of variation found in feedback studies reflects the complexity involved in providing written CF. It also contributes to explaining the large discrepancies among studies and ensuing findings, which in turn prevent L2 practitioners and written CF researchers from obtaining a clearer picture of research findings.

For this reason, future research should consider design recommendations gleaned from this systematic review in order to overcome the current problematic research scenario. In addition to fine-tuning future research efforts, it would be insightful to find ways to tackle issues in key terminology and definitions pertaining to feedback type. Although this study quantified the extent to which a given feedback strategy has been examined (cf. 3.3), inconsistencies in research literature prevent the attainment of any conclusion retarding the number of written CF studies by type (i.e., direct, indirect, or metalinguistic). This requires schematization of written CF to further address comparability issues and advance the L2 error correction practice and research. A study examining this issue is already underway.

References

Abedi, R., Latifi, M., & Moinzadeh, A. (2010). The effect of error correction vs. error detection on Iranian pre-intermediate EFL learners' writing achievement. *English Language Teaching, 3*(4), 168-174. doi: https://doi.org/10.5539/elt.v3n4p168

Ahmadi, D., Maftoon, P., & Mehrdad, A. (2012). Investigating the Effects of Two Types of Feedback on EFL Students' Writing. *Procedia - Social and Behavioral Sciences, 46,* 2590–2595. doi: https://doi.org/10.1016/j.sbspro.2012.05.529

Asassfeh, S. (2013). Corrective Feedback (CF) and English-Major EFL Learners' Ability in Grammatical Error Detection and Correction. *English Language Teaching, 6*(8). doi: https://doi.org/10.5539/elt.v6n8p85

Aveyard, H. (2010). *Doing a Literature Review in Health and Social Care: A Practical Guide.* Maidenhead, UK: McGraw.

Benson, S., & DeKeyser, R. (2018). Effects of written corrective feedback and language aptitude on verb tense accuracy. *Language Teaching Research*, 23(6), 702–726. doi: https://doi.org/10.1177/1362168818770921

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

Bitchener, J. (2009). Measuring the Effectiveness of Written Corrective Feedback: A response to "Overgeneralization from a Narrow Focus: A Response to Bitchener (2008)." *Journal of Second Language Writing*, 18(4), 276–279. doi: https://doi.org/10.1016/j.jslw.2009.06.001

Bitchener, J. & Ferris, D. (2012). *Written Corrective Feedback in Second Language Acquisition and Writing*. UK: Routledge.

Bitchener. (2012). A Reflection on 'The Language Learning Potential' of Written CF. *Journal of Second Language Writing, 21*(4), 348–363. doi: https://doi.org/10.1016/j.jslw.2012.09.006

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

Bitchener, J. & Knoch, U. (2009a). The relative effectiveness of different types of direct written corrective feedback. *System, 37*(2), 322–329. doi: https://doi.org/10.1016/j.system.2008.12.006

Bitchener, J. & Knoch, U. (2009b). The Value of a Focused Approach to Written Corrective Feedback. *ELT Journal, 63*(3), 204–211. doi: https://doi.org/10.1093/elt/ccn043

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. doi: https://doi.org/10.1016/j.jslw.2010.10.002

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. doi: https://doi.org/10.1016/j.jslw.2005.08.001

Bonilla, M., Van Steendam, E., & Buyse, K. (2017). Comprehensive Corrective Feedback on Low and High Proficiency Writers: Examining Attitudes and Preferences. *ITL-International Journal of Applied Linguistics, 168*(1), 91–128. doi: https://doi.org/10.1075/itl.168.1.04bon

Bonilla, M., Van Steendam, E., Speelman, D., & Buyse, K. (2018). The Differential Effects of Comprehensive Feedback Forms in the Second Language Writing Class: Comprehensive Feedback in the L2 Writing Class. *Language Learning, 68*(3), 813–850. doi: https://doi.org/10.1111/lang.12295

Brown, D. (2012). The Written Corrective Feedback Debate: Next Steps for Classroom Teachers and Practitioners. *TESOL Quarterly, 46*(4), 861–867. doi: https://doi.org/10.1002/tesq.63

Bruton, A. (2009a). Designing Research into the Effects of Grammar Correction in L2 Writing: Not So Straightforward. *Journal of Second Language Writing, 18*(2), 136–140. doi: https://doi.org/10.1016/j.jslw.2009.02.005

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

Chandler, J. (2003). The efficacy of various kinds of error feedback for improvement in the accuracy and fluency of L2 student writing. *Journal of Second Language Writing, 12*(3), 267–296. doi: https://doi.org/10.1016/S1060-3743(03)00038-9

Chaudron, C. (1984). The Effects of Feedback on Students' Composition Revisions. *RELC Journal, 15*(2), 1–14.

Cohen, A. (1987). Studying Learner Strategies: How We Get the Information. In A. Wenden & J. Rubin (Eds.), *Learner Strategies in Language Learning* (pp. 31–40). UK: Prentice Hall.

Diab, N. (2015). Effectiveness of written corrective feedback: Does type of error and type of correction matter? *Assessing Writing, 24*, 16–34. doi: https://doi.org/10.1016/j.asw.2015.02.001

El-Ebyary, K., & Windeatt, S. (2010). The Impact of Computer-Based Feedback on Students' Written Work. *International Journal of English Studies, 10*(2), 121–142.

Ellis, R. (2009). A Typology of Written Corrective Feedback Types. *ELT Journal, 63*(2), 97-107. doi: https://doi.org/10.1093/elt/ccn023

Ellis, R. (2010). A Framework for Investigating Oral and Written Corrective Feedback. *Studies in Second Language Acquisition, 32*(02), 335–349. doi: https://doi.org/10.1017/S0272263109990544

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. doi: https://doi.org/10.1016/j.system.2008.02.001

Evans, N., Hartshorn, K., McCollum, R., & Wolfersberger, M. (2010). Contextualizing Corrective Feedback in Second Language Writing Pedagogy. *Language Teaching Research, 14*(4), 445–463. doi: https://doi.org/10.1177/1362168810375367

Evans, N., Hartshorn, K., & Strong-Krause, D. (2011). The Efficacy of Dynamic Written Corrective Feedback for University-Matriculated ESL Learners. *System, 39*(2), 229–239. doi: 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). doi: https://doi.org/10.5430/wje.v2n2p49

Fathman, A. & 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). New York: Cambridge University Press.

Fazio, L. (2001). The effect of corrections and commentaries on the journal writing accuracy of minority-and majority-language students. *Journal of Second Language Writing, 10*(4), 235–249.

Ferris, D. (1997). The Influence of Teacher Commentary on Student Revision. *TESOL Quarterly, 31*(2), 315-339. doi: https://doi.org/10.2307/3588049

Ferris, D. (1999). The Case for Grammar Correction in L2 Writing Classes: A Response to Truscott (1996). *Journal of Second Language Writing, 8*(1), 1–11. doi: https://doi.org/10.1016/S1060-3743(99)80110-6

Ferris, D. (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*(1), 49–62. doi: https://doi.org/10.1016/j.jslw.2004.04.005

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. *Studies in Second Language Acquisition, 32*(02), 181–201. doi: https://doi.org/10.1017/S0272263109990490

Ferris, D., Liu, H., Sinha, A., & Senna, M. (2013). Written Corrective Feedback for Individual L2 Writers. *Journal of Second Language Writing, 22*(3), 307–329. doi: https://doi.org/10.1016/j.jslw.2012.09.009

Ferris, D., & Roberts, B. (2001). Error Feedback in L2 Writing Classes: How Explicit Does it Need to Be? *Journal of Second Language Writing, 10*(3), 161–184.

Frantzen, D. (1995). The effects of grammar supplementation on written accuracy in an intermediate Spanish content course. *The Modern Language Journal, 79*(3), 329–344.

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

Gaskell, D., & Cobb, T. (2004). Can Learners Use Concordance Feedback for Writing Errors? *System, 32*(3), 301–319. doi: https://doi.org/10.1016/j.system.2004.04.001

Goldstein, L. (2004). Questions and answers about teacher written commentary and student revision: teachers and students working together. *Journal of Second Language Writing, 13*(1), 63–80. doi: https://doi.org/10.1016/j.jslw.2004.04.006

Guénette, D. (2007). Is Feedback Pedagogically Correct? *Journal of Second Language Writing, 16*(1), 40–53. doi: https://doi.org/10.1016/j.jslw.2007.01.001

Guénette, D. (2012). The Pedagogy of Error Correction: Surviving the Written Corrective Feedback Challenge. *TESL Canada Journal, 30*(1), 117-126.

Hanaoka, O., & Izumi, S. (2012). Noticing and Uptake: Addressing Pre-articulated Covert Problems in L2 Writing. *Journal of Second Language Writing, 21*(4), 332–347. doi: https://doi.org/10.1016/j.jslw.2012.09.008

Hartshorn, K., Evans, N., Merrill, P., Sudweeks, R., Strong-Krause, D., & Anderson, N. (2010). Effects of Dynamic Corrective Feedback on ESL Writing Accuracy. *TESOL Quarterly, 44*(1), 84-109. doi: https://doi.org/10.5054/tq.2010.213781

Hedgcock, J., & Lefkowitz, N. (1992). Collaborative Oral/Aural Revision in Foreign Language Writing Instruction. *Journal of Second Language Writing, 1*(3), 255–276.

Hendrickson, J. (1978). Error Correction in Foreign Language Teaching: Recent Theory, Research, and Practice. *The Modern Language Journal, 62*(8), 387–398.

Hyland, K., & Hyland, F. (2006). Feedback on Second Language Students' Writing. *Language Teaching, 39*(02), 83-101. https://doi.org/10.1017/S0261444806003399

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. doi: https://doi.org/10.1111/modl.12189

Kepner, C. (1991). An Experiment in the Relationship of Types of Written Feedback to the Development of Second-Language Writing Skills. *The Modern Language Journal, 75*(3), 305–313.

Lalande, J. (1982). Reducing Composition Errors: An Experiment. *The Modern Language Journal, 66*(2), 140–149.

Lavolette, E., Polio, C., & Kahng, J. (2015). The Accuracy of Computer-Assisted Feedback and Students' Responses to It. *Language, Learning & Technology, 19*(2), 50-68. Retrieved from https://bit.ly/2BlItLL

Lee, I. (1997). ESL Learners' Performance in Error Correction in Writing: Some Implications for Teaching. *System, 25*(4), 465–477.

Lee, I. (2014). Feedback in Writing: Issues and Challenges. *Assessing Writing, 19*, 1–5. doi: https://doi.org/10.1016/j.asw.2013.11.009

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. doi: https://doi.org/10.1016/j.jslw.2015.08.011

MacDonald, R. (1991). Developmental students' processing of teacher feedback in composition instruction. *Review of Research in Developmental Education, 8*(5), 1-8.

Manchón, R. (2011). *Learning-to-write and Writing-to-Learn in an Additional Language.* Amsterdam: John Benjamins.

Martínez, E., & De Larios, R. (2010). The Use of Models as a Form of Written Feedback to Secondary School Pupils of English. *International Journal of English Studies, 10*(2), 143–170.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G. & The Prisma Group. (2009). Preferred Reporting Items for Systematic Reviews and Meta-analyses: The PRISMA statement. PLoS Medicine, .(7), e1000097. doi: http://dx.doi.org/10.1371/journal.pmed.1000097

Murphy, L., & De Larios, R. (2010). Feedback in Second Language Writing: An Introduction. *International Journal of English Studies, 10*(2), 1-15.

Nagata, N. (1997). The Effectiveness of Computer-Assisted Metalinguistic Instruction: A Case Study in Japanese. *Foreign Language Annals, 30*(2), 187–200.

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. doi: https://doi.org/10.1016/j.jslw.2012.09.004

Qi, D., & Lapkin, S. (2001). Exploring the Role of Noticing in a Three-Stage Second Language Writing Task. *Journal of Second Language Writing, 10*(4), 277–303.

Russell, J., & Spada, N. (2006). The Effectiveness of Corrective Feedback for the Acquisition of L2 Grammar. In J. M. Norris & L. Ortega (Eds.), *Synthesizing Research on Language Learning and Teaching* (pp. 133–164). John Benjamins Publishing Company.

Sampson, A. (2012). "Coded and Uncoded Error Feedback: Effects on Error Frequencies in Adult Colombian EFL learners Writing". *System, 40*(4), 494–504. doi: https://doi.org/10.1016/j.system.2012.10.001

Santos, M., López-Serrano, S., & Manchón, R. (2010). The differential effect of two types of direct written corrective feedback on noticing and uptake: Reformulation vs. error correction. *IJES, International Journal of English Studies, 10*(1), 131–154.

Semke, H. (1984). Effects of the Red Pen. *Foreign Language Annals*, 17(3), 195–202.

Sheen, Y. (2007). The Effect of Focused Written Corrective Feedback and Language Aptitude on ESL Learners' Acquisition of Articles. *TESOL Quarterly, 41*(2), 255–283. doi: https://doi.org/10.2307/40264353

Sheen, Y. (2010). Differential Effects of Oral and Written Corrective Feedback in the ESL Classroom. *Studies in Second Language Acquisition, 32*(02), 203–234. https://doi.org/10.1017/S0272263109990507

Sheppard, K. (1992). Two Feedback Types: Do They Make A Difference? *RELC Journal, 23*(1), 103–110. https://doi.org/10.1177/003368829202300107

Shintani, N. (2015). The effects of computer-mediated synchronous and asynchronous direct corrective feedback on writing: a case study. *Computer Assisted Language Learning*, 29(3), 517-538. doi: https://doi.org/10.1080/09588221.2014.993400

Shintani, N. & Aubrey, S. (2016). The Effectiveness of Synchronous and Asynchronous Written Corrective Feedback on Grammatical Accuracy in a Computer-Mediated Environment. *The Modern Language Journal, 100*(1), 296–319. doi: https://doi.org/10.1111/modl.12317

Shintani, & Ellis, R. (2013). The comparative effect of direct written corrective feedback and metalinguistic explanation on learners' explicit and implicit knowledge of the English indefinite article. *Journal of Second Language Writing, 22*(3), 286–306. doi: https://doi.org/10.1016/j.jslw.2013.03.011

Shintani, N, & Ellis, R. (2015). Does language analytical ability mediate the effect of written feedback on grammatical accuracy in second language writing? *System, 49*, 110–119. https://doi.org/10.1016/j.system.2015.01.006

Shintani, N., Ellis, R., & Suzuki, W. (2014). Effects of Written Feedback and Revision on Learners' Accuracy in Using Two English Grammatical Structures: Effects of Written Feedback and Revision. *Language Learning, 64*(1), 103–131. doi: https://doi.org/10.1111/lang.12029

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. doi: https://doi.org/10.1111/modl.12212

Stiff, R. (1967). The Effect upon Student Composition of Particular Correction Techniques. *Research in the Teaching of English, 1*(1), 54–75.

Storch, N. (2010). Critical Feedback on Written Corrective Feedback Research. *International Journal of English Studies, 10*(2), 29–46.

Storch, N., & Wigglesworth, G. (2010). Learners' Processing, Uptake and Retention of Corrective Feedback on Writing. *Studies in Second Language Acquisition, 32*(02), 303–334. doi: https://doi.org/10.1017/S0272263109990532

Truscott, J. (1996). The Case Against Grammar Correction in L2 Writing Classes. *Language Learning, 46*(2), 327–369.

Truscott, J. (2001). Selecting Errors for Selective Error Correction. *Concentric: Studies in English Literature and Linguistics, 27*(2), 93–108.

Truscott, J. (2004). Evidence and Conjecture on the Effects of Correction: A Response to Chandler. *Journal of Second Language Writing, 13*(4), 337–343. doi: https://doi.org/10.1016/j.jslw.2004.05.002

Truscott, J. (2007). The Effect of Error Correction on Learners' Ability to Write Accurately. *Journal of Second Language Writing, 16*(4), 255–272. doi: 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. doi: https://doi.org/10.1016/j.jslw.2008.05.003

Van Beuningen, C. (2010). Corrective feedback in L2 writing: Theoretical Perspectives, Empirical Insights, and Future Directions. *International Journal of English Studies, 10*(2), 1–27.

Van Beuningen, C., De Jong, N., & 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., De Jong, N., & Kuiken, F. (2012). Evidence on the Effectiveness of Comprehensive Error Correction in Second Language Writing: Effectiveness of Comprehensive CF. *Language Learning, 62*(1), 1–41. doi: https://doi.org/10.1111/j.1467-9922.2011.00674.x