

An analysis of pedagogical and theoretical implications: Lessons from studies with English as a foreign language learners

Análisis de implicaciones teóricas y pedagógicas: lecciones de estudios con aprendices de inglés como una lengua extranjera

MARISELA BONILLA LÓPEZ

Escuela de Lenguas Modernas, Universidad de Costa Rica

marisela.Bonilla@ucr.ac.cr

ORCID ID: [0000-0002-1194-7721](https://orcid.org/0000-0002-1194-7721)

Abstract

Different researchers in the field of error correction have voiced their concern and advocated for research that is conducted in under-represented settings (e.g., Lee, 2014), that investigates a feedback scope in line with common second language (L2) classroom feedback practices (e.g., Hartshorn et al., 2010), and that looks into factors beyond the end product such as learner variables (e.g., Storch & Wigglesworth, 2010). Against this background, an analysis of previous literature in light of said concerns is called for. Specifically, notwithstanding the valuable contribution of past feedback studies (e.g., Benson & DeKeyser, 2018; Karim & Nassaji, 2020; Kurzer, 2022), three in particular (i.e., Bonilla et al., 2017, 2018, 2021) are worth further discussing for various reasons. The purpose of this paper is threefold: (1) to offer an overview of the design and main findings of the studies, (2) to critically analyze their theoretical contribution to the literature on written CF, and (3) to sketch the practical application of their findings to the L2 writing class.

Key words: error correction, written CF, second language classroom, theoretical and practical implications

Resumen

Diferentes investigadores en el campo de la corrección de errores han expresado su preocupación y han abogado por la investigación que se lleva a cabo en entornos subrepresentados (p. ej., Lee, 2014), que investiga un alcance de retroalimentación en línea con las prácticas comunes de retroalimentación en el aula de la segunda lengua (L2) (p. ej., Hartshorn et al., 2010), y que analiza factores más allá del producto final, como las variables del alumno (p. ej., Storch y Wigglesworth, 2010). En este contexto, se requiere un análisis de la literatura previa a la luz de dichas preocupaciones. Específicamente, a pesar de la valiosa contribución de estudios de retroalimentación anteriores (p. ej., Benson y DeKeyser, 2018; Karim y Nassaji, 2020; Kurzer, 2022), tres en particular (es decir, Bonilla et al., 2017, 2018, 2021) requieren un análisis más profundo por varias razones. El propósito de este artículo es triple: (1) ofrecer una visión general del diseño y los principales hallazgos de los estudios, (2) discutir críticamente su contribución teórica a la literatura sobre FC escrita, y (3) esbozar la aplicación de sus hallazgos a la clase de escritura L2.

Palabras clave: corrección de errores, realimentación correctiva escrita, aula de una segunda lengua, implicaciones teóricas y prácticas

Introduction

Over the years, the error correction practice has been heavily contested (e.g., Truscott, 1996, 1999, 2001; Truscott & Hsu, 2008). However, such criticism only prompted second language (L2) researchers to conduct more research on written corrective feedback (CF) and L2 composition teachers to continue embracing a practice that has been ubiquitous to this day. Indeed, research on written CF has come a long way. During the process, mixed or inconclusive findings have been obtained due to differences in methodologies, measures to assess the feedback effect, and/or design and execution shortcomings (cf. Ferris, 2004; Gu enette, 2007; K. Hyland & F. Hyland, 2006; Storch, 2010). However, it could be said that after over 40 years of research on written CF and given the most recent empirical contribution of second language acquisition (SLA)-oriented feedback studies, the debate is no longer whether written CF is effective. The *new* debate is the extent to which such findings hold in other instructional contexts when different written CF strategies are implemented, other linguistic features targeted, and other variables examined.

In addition, this research scenario becomes more complex when even with a significant amount of compelling empirical evidence (e.g., van Beuningen et al., 2012), L2 writing practitioners are still left with insufficient practical suggestions for use in the L2 classroom. That is why different researchers have voiced their concern and advocated for research that is conducted in under-represented settings (e.g., Lee, 2014), that investigates a feedback scope in

line with common L2 classroom feedback practices (e.g., Hartshorn et al., 2010), and that has a design that looks into factors beyond the end product such as learner variables (e.g., Storch & Wigglesworth, 2010).

Against this background, an analysis of previous literature in light of said concerns is called for. Specifically, notwithstanding the valuable contribution of past feedback studies (e.g., Benson & DeKeyser, 2018; Karim & Nassaji, 2020; Kurzer, 2022), three in particular (i.e., Bonilla et al., 2017, 2018, 2021) are worth further discussing for a number of reasons. First, these are the only series of studies that employed the same (largely unexplored) learner type (i.e., English majors in a FL setting). Second, all three studies examined a feedback practice that is very common in FL classrooms—yet highly criticized in part of the literature (i.e., correcting a large array of errors at once). Third, they also looked into response variables beyond grammatical accuracy. For these reasons, the purpose of this paper is threefold: (1) to offer an overview of the design and main findings of the studies, (2) to critically discuss their theoretical contribution to the literature on written CF, and (3) to sketch the actual practical application of their findings to the L2 writing class.

An Overview of the Studies

Table 1 summarizes the design of Bonilla et al. (2017, 2018, 2021). As can be seen, in their attempt to determine the effect of feedback on overall grammatical and non-grammatical accuracy (Bonilla et al., 2017, 2018) and

individual linguistic categories (Bonilla et al., 2021), the researchers focused on feedback practices that have been overlooked in research designs despite being common in EFL settings (van Beuningen et al., 2012). Namely, the authors targeted a large array of errors (i.e., comprehensive CF) and employed two main feedback types: direct corrections (i.e., provision of the correct target language form above the error) and metalinguistic CF (i.e., rule reminders or codes). From the table, it is also worth highlighting that the studies answered to calls for data on learner factors (e.g., cognitive load, preferences, and attitudes) and took place in a non-English dominant setting—as opposed to the large majority of SLA oriented studies.

In a nutshell, as far as accuracy is concerned, statistical analyses revealed that providing written CF is better than no feedback at all. That is, in all three studies there were significant differences between the experimental groups and control after revision and in posttests. Nonetheless, no significant differences were found between experimental groups, meaning that under certain conditions, both feedback strategies (direct corrections and metalinguistic feedback either with reminders or codes) are equally effective to enhance learners' grammatical accuracy in the short and long term. This was not the case with groups of the same feedback type but different feedback scope as in Bonilla et al. (2018, 2021).

Table 1
Summary of the Experimental Studies

	Bonilla et al. (2017)	Bonilla et al. (2018)	Bonilla et al. (2021)
Macro context	Costa Rican EFL majors	Costa Rican EFL majors	Costa Rican EFL majors
Sample	First-year majors (N = 52) Fourth-year majors (N = 39)	First-year majors (N = 139)	First-year majors (N = 139)
Proficiency level	Elementary Advanced	Low intermediate	Low intermediate
Feedback strategies	Direct corrections (DCF) Metalinguistic rules (MER)	Direct corrections (DCF) Metalinguistic codes (MEC)	Direct corrections (DCF) Metalinguistic codes (MEC)

Treatment and control	Low proficiency	DCF on grammatical issues ($n = 29$)	DCF on grammatical issues ($n = 29$)
	DCF ($n = 18$)	MEC on grammatical issues ($n = 28$)	MEC on grammatical issues ($n = 28$)
	MER ($n = 17$)	DCF on grammatical and non-grammatical issues ($n = 27$)	DCF on grammatical and non-grammatical issues ($n = 27$)
	Control ($n = 17$)	MEC on grammatical and non-grammatical issues ($n = 28$)	MEC on grammatical and non-grammatical issues ($n = 28$)
	High proficiency	Control ($n = 27$)	Control ($n = 27$)
	DCF ($n = 14$)	Comprehensive (13 error types)	Comprehensive (13 error types)
Feedback scope	MER ($n = 12$)		
	Control ($n = 13$)		
Duration	8 weeks	6 weeks	6 weeks
Data collection tools			
Accuracy	Argumentative texts	Argumentative texts	Argumentative texts
Attitudes	Questionnaire / Interview	Questionnaire	X
Preferences	Questionnaire / Interview	X	X
Cognitive load	X	Mental effort scale	X
Statistical tests	Two-way ANOVA	Mixed-effect linear model	Mixed-effect linear model
	Two-way ANCOVA		Logistic regression
	Chi-square		

Note: Elaborated by author.

Main findings from Bonilla et al. (2018, 2021) showed that groups with differing numbers of targeted error categories did have significant statistical differences for grammatical errors. For example, learners whose attention was targeted at grammatical issues only were able to significantly improve their grammatical accuracy more than those whose attentional resources were directed at both grammatical and non-grammatical errors.

As for learner variables, Bonilla et al. (2017, 2018) showed that while learners' attitudinal response to a given treatment may be favorable as a whole, significant differences may take place depending on the construct being measured. To illustrate, the participants in Bonilla et al. (2017) rendered statistically significant differences per level (that is, low proficiency versus high proficiency), namely feedback appropriateness. Bonilla et al. (2017)

found significant differences per treatment, where those correcting errors with codes reported understanding the feedback significantly less than those correcting the same error types with direct corrections. Furthermore, main findings in Bonilla et al. (2017) revealed that feedback preferences may also hinge on learners' level, being high proficiency level learners more likely to prefer metalinguistic feedback than their low proficiency counterparts. Of equal relevance are the significant differences found in learners' report of cognitive load in Bonilla et al. (2018), which provide evidence of some feedback treatments being more overburdening than others. For instance, correcting grammatical errors with direct corrections had significantly lower reports of cognitive load than correcting grammatical and non-grammatical errors with metalinguistic codes.

Contribution to Error Correction Research

It could be argued that Bonilla et al. (2017, 2018, 2021) provide theoretical knowledge in four main ways:

Reconciling L2 Writing and SLA Research

Bonilla et al., (2017, 2018, 2021) bridge the gap between L2 writing and SLA feedback research in two main ways: (1) they adopt the design of L2 acquisition feedback studies (i.e., pretest-posttest-posttest) besides incorporating a revision component and (2) they frame the findings within a theory of SLA, namely a cognitive/interactionist perspective.

The Research Design

A number of researchers have pointed out the need to reconcile L2 writing and SLA feedback research strands and suggested that one way to do so is through the research design (e.g., Ferris, 2010; Sheen, 2010). Specifically, they stress the need to adopt the methodology that is used in oral CF studies (e.g., Lyster, 2004; Sheen, 2004). The importance of such a suggestion is that a pretest-posttest-delayed posttest design looks into what ultimately matters from the L2 acquisition standpoint (i.e., L2 development) while addressing a measure that is relevant for L2 composition teachers (i.e., accuracy). For example, Sheen (2010) states that “inquiry into written CF within the SLA research paradigm can be seen as relevant to L2 writing pedagogy, given that one of the aims of such pedagogy is to improve students' written grammatical accuracy” (p. 211). Similar to Sheen (2010), Ferris (2010) refers to the importance of adopting the methodology of oral CF studies but highlights that incorporating the revision component is necessary as well. To Ferris (2010), such a design, which the author labels as “blended” (p. 195), bridges the gap between two lines of inquiry that have looked into the same phenomenon but with two differing starting points. In fact, those differences are reflected not only in researchers' empirical interests but also in teachers' reasons to provide written CF. As Bitchener (2012b) explains, “[w]hile composition teachers may be more likely to do this so that their learners can edit their writing and produce error-free revisions, language learning teachers may do so in order to help their learners acquire

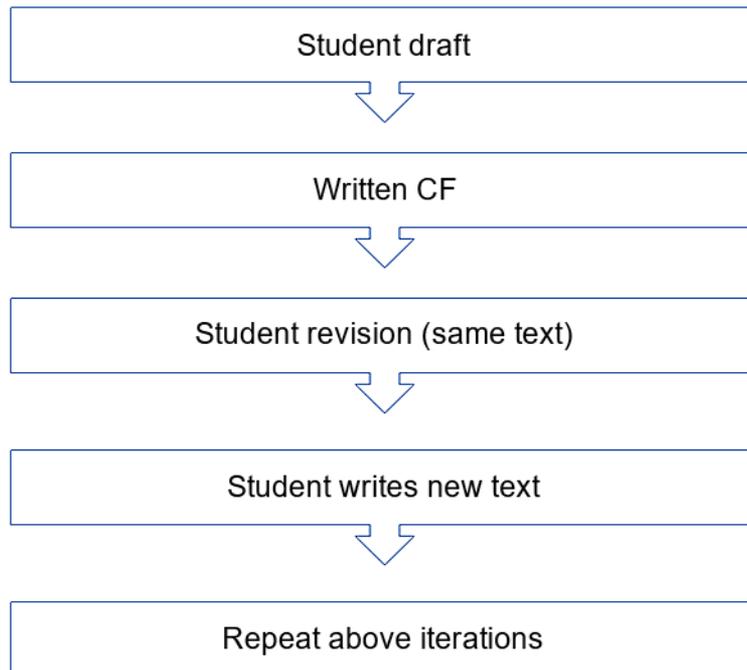
specific target-like forms and structures, demonstrated in the writing of new texts” (Bitchener, 2012b, p. 855).

Consequently, with previous design-related suggestions in mind (e.g., Ferris, 2010; Sheen, 2010), learners in Bonilla et al. (2017, 2018, 2021) wrote an initial draft, received feedback on it, revised the text, and produced a new one. Figure 1 illustrates the stages

needed in a blended design study to examine the effectiveness of written CF. By following the stages in Figure 1, Bonilla et al. (2017, 2018, 2021) are a valuable addition to the literature as they delve into the value of written CF both as an instructional intervention to help learners successfully edit their texts and as a learning tool to promote L2 development.

Figure 1

A blended design to examine written CF effectiveness (Ferris, 2010, p. 195)



The Interpretation of Results

After decisions about the research design have been made and results have been obtained, two questions worth pondering upon relate to what the findings mean and which SLA theory (i.e., theoretical perspective) could be used to interpret those findings. If these questions are addressed, such an attempt could be construed as yet another way of both reconciling the L2 writing and SLA lines of inquiry and in particular moving the field of L2 writing studies forward (cf. Polio, 2012)—an element present in Bonilla et al. (2017, 2018, 2021). In the case of the findings from Bonilla et al. (2017, 2018, 2021), they were interpreted from a cognitive/interactionist standpoint since it is the one that “has arguably the most to offer” (Bitchener, 2012a, p. 351). Indeed, despite being originally intended for oral production, researchers acknowledge that the cognitive/interactionist perspective is the one that has the most room to explain the likely effects of written CF and its L2 learning potential (Bitchener, 2012a; Bitchener & Ferris, 2012; Polio, 2012). For example, applied to L2 writing, Bitchener (2012a) explains that an interactionist standpoint (1) acknowledges the crucial role that input (e.g., negative evidence such as written CF), noticing (e.g., studying the feedback), and output (e.g., a revised or a new text) play in SLA processes; (2) is clear in that something more than mere L2 exposure is needed for L2 development, hence, the relevance of ‘pushed’ output for producing modified output; and (3) stresses the importance of attention in facilitating L2 learning (e.g., Schmidt, 1990, 2001).

Therefore, interpreted from the aforementioned perspective, the results obtained in Bonilla et al., (2017, 2018, 2021) are welcoming news for both SLA and L2 writing researchers and teachers. Overall, the findings mean that the time and effort spent providing corrections (comprehensively and via direct CF or metalinguistic CF) is not in vain: learners (such as those in the three studies) can focus on a wide range of errors, they can notice the mismatch between the input and their output, process the corrections, learn from them, and apply what they learn in newly produced texts. In other words, learners can follow the stages of cognitive processing of input (see Gass, 1997 cited in Bitchener, 2016).

Yielding Evidence on the Effects of Comprehensive CF

The theoretical contribution and implications of Bonilla et al., (2017, 2018, 2021) can be discussed in terms of the accuracy, cognitive, and attitudinal effects of comprehensive CF.

Accuracy Effect

To analyze the relevance of the accuracy results obtained in Bonilla et al., (2017, 2018, 2021), a basic guide can be gleaned in Truscott's (1996) highly contested article. The author summarizes his logic behind result interpretation when groups that receive written CF are compared with those that do not. According to Truscott (1996), accuracy results could be viewed as follows:

- a. If learners that have error correction perform just as those that have none, “correction is not helpful” (p. 329).

- b. If learners that have no error correction perform better than those that have it, "correction is apparently harmful" (p.329).
- c. If learners that have error correction perform significantly better than those that have none, "correction is important for learning" (p.329).

While results (a) and (b) may not exactly warrant such conclusions (for a rebuttal, see Bruton, 2010; Ferris, 1999, 2004), Truscott's (1996) summary serves to show that even through the lens of the most vocal opponent of written CF, option (c) means that the findings in Bonilla et al., (2017, 2018, 2021) constitute evidence of L2 learning: all studies had experimental groups that were able to significantly outperform the control group not only in a revised draft but also in new texts, being the latter the results that matter for Truscott (1996) as far as L2 learning evidence is concerned.

On the whole, the findings in Bonilla et al., (2017, 2018, 2021) are worth noting because they run counter to claims about comprehensive CF being "ineffective" (Ellis et al., 2008, p. 368) and unlikely to lead to L2 development due to processing issues (cf. Bitchener, 2008; Sheen, 2007). This begs the question of what could make a given group of L2 learners able to process a large array of errors without it being too much to handle as the literature suggests (e.g., Bitchener, 2008; Ellis et al., 2008; Sheen et al., 2009). The answer may lie in the offline nature of writing, which gives learners ample opportunities to pay attention to the feedback, engage in a cognitive comparison, and modify their output (Bitchener, 2012a; Polio, 2012; Sheen, 2010; van Beuningen, 2010).

The relevance of such an interpretation is that it not only acknowledges the differences between oral CF and written CF but also brings into question claims about learners' attentional capacity being limited to attend to comprehensive CF. For instance, the fact that learners may lack attentional resources to process comprehensive CF is grounded on models on the role of attention in L2 acquisition (e.g., Skehan, 1998, 2003, Robinson, 2005), yet such models "were less concerned with understanding writing itself or with investigating language learning through writing" (Byrnes & Manchón, 2014, p. 4). Thus, the accuracy improvement in Bonilla et al., (2017, 2018, 2021) add to the incipient knowledge base on the L2 learning potential of comprehensive CF (see also van Beuningen et al., 2008, 2012); they also lend support to Polio (2012), who argues that the jury is out when asserting that learners' attentional resources could be diverted when attending to written CF.

Cognitive Effect

Before Bonilla et al. (2018), to the researcher's knowledge there was no empirical evidence to validate the assertion that comprehensive CF is cognitively overloading to the extent that learners cannot process it (e.g., Ellis et al., 2008; Sheen et al., 2009). That is, no actual research attempt had been made to measure the cognitive load of processing comprehensive CF. That is why if there was any truth to this claim, it was reasonable to assume that learners' estimates of cognitive load (i.e., the mental effort ratings) after revision with comprehensive CF forms would show.

Interestingly, the results obtained (albeit the small dent) were mostly in line with what had been hypothesized: that some conditions would be more cognitively complex and that, as a result, they would report higher cognitive load estimates (see page 832 in Bonilla et al., 2018). However, the findings also hint at a scenario not considered before in the literature (e.g., Ellis et al. 2008): that the overloading nature of a feedback treatment may not solely rest on its broad scope but also on a low degree of explicitness. Such a conclusion originates from the results that show that comprehensive direct feedback forms proved significantly less overloading than comprehensive metalinguistic ones and that direct corrections on grammatical errors imposed the lowest cognitive load. On the whole, the significant cognitive load differences between comprehensive groups could serve as support to claims that indeed correcting a large array of errors may be overburdening for learners (cf. Bitchener & Ferris, 2012; Evans et al., 2010; McMartin-Miller, 2014), yet that same fact could take away part of the stigma surrounding comprehensive CF over for years: an overloading effect may not exclusively be the result of a broad approach to errors but of a comprehensive CF practice in combination with feedback strategies of differing explicitness.

Attitudinal Effect

The studies add a further dimension to current understanding of learners' affective response to written CF generally and comprehensive CF particularly by establishing comparisons

by level (Bonilla et al., 2017) and condition (Bonilla et al., 2018). To start, it may be a common belief that in general students find feedback helpful even when there is some frustration—especially if one considers Hedgcock and Lefkowitz's (1994) descriptive study. However, as these same authors assert, “learners' perception of what constitutes useful feedback vary according to the orientation and demands of the L2 learning context” (p. 157). With this in mind, caution should be exercised when attempting to apply previous findings (be it from the large number of descriptive studies or the few empirical feedback ones) to an under-researched instructional context such as that of Bonilla et al. (2017, 2018). The attitudinal variability that can be present in any given instructional context is shown in Table 2.

Table 2
Overview of affective response to written CF

Study	Type	Setting	Response to CF
Bonilla et al. (2017, 2018)	Empirical	English (Teaching) majors in Costa Rica	Favorable without signs of frustration
Hyland (2001)	Descriptive	Distance EFL in Hong Kong	Favorable without signs of frustration
Hedgcock and Lefkowitz (1994)	Descriptive	FL and SL learners in the United States	Favorable with signs of frustration
Lee (2008)	Descriptive	EFL low proficiency learners in Hong Kong	Unfavorable with signs of frustration
Sampson (2012)	Empirical	EFL learners in Colombia	Unfavorable with signs of frustration

Source: Elaborated by author

Why then does exploration of learners' attitudes matter? They do for two reasons: (1) Resistance to feedback may lead to lack of uptake (Storch & Wigglesworth, 2010; Swain, 2006; Swain & Lapkin, 2002). As pointed out in Bitchener and Storch (2016), "if learners are given a particular type of written CF and do not believe it is helpful, they may decide to ignore it, that is, not to attend to it and not cognitively process it" (p. 30). In the case of Bonilla et al. (2017, 2018), there is no evidence that learners' affective response to feedback was unwelcoming to the extent that L2 development could not take place. (2) Certainly, the picture painted in Bonilla et al. (2017, 2018) does not match a potential assumption that in the end L2 learners' perception of usefulness may be the same (with or without signs of frustration). As a matter of fact, the results for attitudes towards written CF reveal a more intricate reality than

that: the utility of a given treatment is just one of other constructs that learners may react to favorably, being others not as positively perceived—hence, the significant level differences for feedback appropriateness (see Bonilla et al., 2017) and the significant condition differences for feedback comprehensibility (see Bonilla et al., 2018). Therefore, while adding a new variable to the equation (i.e., feedback explicitness), the results are also a confirmation of the fact that what could make a difference in learners' reactions to written CF are learner (Hyland, 1998) and contextual (Hedgcock & Lefkowitz, 1996) variables.

Delving into the Differential Effect of Direct and Metalinguistic CF

By employing direct corrections and two forms of metalinguistic CF (i.e., rule reminders and codes), Bonilla et al. (2017, 2018, 2021) were able

to explore to what extent the degree of explicitness of the feedback strategies may play a role in their effectiveness. After all, such a variable had already been suggested as paramount. For example, Sheen (2010) affirmed that “the crucial factor that influences the effectiveness of CF is the explicitness of the feedback (i.e., whether its corrective force is clear)” (p. 225). However, prior to Bonilla et al. (2017, 2018, 2021), there had not been a series of studies that had been conducted within the same contextual setting (i.e., EFL), with the same learner type (i.e., English majors), and the same feedback types (i.e., direct corrections and metalinguistic CF). The relevance of these design similarities across studies is

that they allow to reach firmer conclusions about the effectiveness of feedback, especially because notwithstanding their great contribution, design differences in past studies have not allowed comparisons.

Against this backdrop, when analyzing what the written CF research base has posited about each feedback strategy and what the (qualitative and quantitative) data from Bonilla et al. (2017, 2018, 2021) actually rendered, it is possible to glean their theoretical contribution to current knowledge on the differential effect of direct CF and metalinguistic CF. Table 3 provides an overview of previous claims which Bonilla et al., (2017, 2018, 2021) corroborate through their findings.

Table 3
Overview of Claims that Find Support in the Experimental Studies

Feedback strategy	Bonilla et al. (2017)	Bonilla et al. (2018)	Bonilla et al. (2021)
Direct corrections			
Lessen confusion from lack of understanding of metalinguistic codes (Bitchener & Knoch, 2008; Ferris & Roberts, 2001)		√	
Provide explicit guidance about how to correct even the most complex errors (Ellis, 2009b, Ellis & Shintani, 2014)			√
Are facilitative of L2 learning (Bitchener, 2012a)	√	√	√
Metalinguistic codes			
Pose comprehensibility issues (Hedgcock & Lefkowitz, 1996; Ferris, 1995)		√	
Afford opportunities for guided learning, which contribute to L2 development (Lalande, 1982)		√	
Metalinguistic reminders			
Their deep level of processing may be facilitative of L2 learning (Ellis, 2010)	√		

Source: Elaborated by author

Addressing what EFL Writers Bring to the Feedback Experience

The last theoretical contribution of the analyzed studies is their insight into the role of learner variables such as preferences (Bonilla et al., 2017) and proficiency level as well as attitudes (Bonilla et al., 2018) in learners' engagement with feedback. Their findings clearly add to the research base by adding support for Oladejo's (1993) assumption that "with increasing levels of competence, the preferences and wants of learners will change" (p. 74) and to Enginarlar's (1993) observation of his descriptive findings where more advanced students "may have considered the procedure as very didactic but not extremely necessary for themselves" (p. 197).

The implications of the existence of an association between proficiency level and both learners' preferences and attitudes towards a written CF treatment are noteworthy. For instance, the evidence brings into question one-size-fits-all approaches to error correction where learners' inaccuracies are treated in the same way irrespective of their L2 proficiency and where feedback practices remain (completely) oblivious to L2 learners' preferences. Because "learner-internal factors, such as those with a motivational and affective component ... operate on an emotional level and influence whether a learner initiates language learning processes" (Bitchener & Storch, 2016, p. 29), overlooking such variables could ultimately be detrimental for L2 development.

To summarize, that proficiency level could play a role on variables at a motivational/affective level puts a number of other relevant theoretical and pedagogical aspects to the table.

If the application of findings from feedback studies is to be made and replication studies are to be conducted, learners' proficiency level needs to be carefully measured and reported (see Gu enette, 2007). In addition, the possibility that (1) low or high proficient EFL writers bring instrumental goals and self-perceptions of L2 proficiency to the feedback experience and that (2) these factors may impact learners' feedback preferences and attitudes, makes a stronger case for the following: understanding how they view themselves as L2 learners, eliciting why they are learning a L2, finding out their feedback preferences, and probing their attitudes after written CF treatment.

Lessons to Be Learned

On the basis of the findings in Bonilla et al. (2017, 2018, 2021), this subsection will briefly present the general conclusions and ensuing pedagogical suggestions for the L2 writing class. They will specifically deal with four main areas: input, feedback strategies, accuracy, cognitive load, and learner factors.

Input

Conclusion 1: Written CF may not be a waste of L2 teachers' time and effort. The evidence obtained in Bonilla et al. (2017, 2018, 2021) is a testament to the effectiveness of written CF. Similar to more recent pretest-posttest-delayed posttest studies (e.g., Bitchener & Knoch, 2010a; Diab 2015), these studies demonstrate that more than being "a dramatic failure" (Truscott, 2007, p. 271), written CF proved to be an effective pedagogical intervention

that brought about an increase of (grammatical and non-grammatical accuracy) in the short- and long-term.

Pedagogical suggestion: Correct learners' written errors confidently.

Conclusion 2: Comprehensive CF may not be detrimental to learners' feedback processing abilities. Contrary to claims against comprehensive CF (e.g., Bitchener, 2008; Sheen et al., 2009), none of the studies analyzed found evidence that learners' attention to multiple errors was overwhelming to the extent that they could not attend to the feedback, learn from it, and retrieve it to apply in new contexts. This clears any doubts about learners' inability to process a large array of errors and confirms the language learning potential of comprehensive CF (e.g., van Beuningen et al., 2012).

Pedagogical suggestion: Do not discard comprehensive CF as a worthwhile error correction practice.

Feedback Strategies

Conclusion 3: Both direct corrections and metalinguistic reminders (rules) are effective feedback strategies to enhance learners' editing skills and to improve their language use over time.

Bonilla et al. (2017) found that irrespective of proficiency level (e.g., low or high), both direct corrections and metalinguistic reminders yielded significant grammatical accuracy gains in the short and long term. This opens more doors for L2 (composition) teachers when it comes to improving grammatical accuracy: if time, workload, or even lack of extensive linguistic

knowledge prevents teachers from providing clear, accurate rules, then an equally useful strategy such as direct corrections is at their disposal.

Pedagogical suggestion: Employ either direct corrections or metalinguistic reminders (rules) for short- or long-term grammatical accuracy changes.

Conclusion 4: The effect of direct corrections on overall accuracy may be superior to that of metalinguistic codes.

The findings in Bonilla et al. (2018) showed that both direct corrections and metalinguistic codes brought about successful error correction of grammatical and non-grammatical errors. However, contrary to direct corrections, metalinguistic codes lost any advantage they had on non-grammatical accuracy in text revision. Also, for grammatical development, metalinguistic codes did not prove any more advantageous than other experimental groups whereas direct corrections did.

Pedagogical suggestion: Use either direct corrections or metalinguistic codes if the aim is enhancing short-term overall grammatical and non-grammatical accuracy. Nonetheless, if the feedback purpose is L2 development, opt for direct corrections.

Conclusion 5: Direct corrections could be more effective in enabling low intermediate learners to correct morphological, syntactic, and stylistic errors.

Bonilla et al. (2021) demonstrated that direct corrections were vastly superior to its metalinguistic counterparts. The findings showed that metalinguistic codes only had a significant effect on one error category

(i.e., subject-verb agreement) and for one measurement only (i.e., correction success during revision). For all other categories where significance was found, direct corrections were the only ones that brought about an immediate (e.g., on pronoun and subject deletion errors) and a lasting effect (e.g., on article, prepositions, sentence structure, and word form errors).

Pedagogical suggestion: To increase the likelihood of successful error treatability of individual (grammatical and non-grammatical) targets, treat them with direct corrections.

Conclusion 6: Direct corrections can be implemented in ways that learners' background knowledge of the targeted features can be activated (e.g., studying the feedback) and problem-solving opportunities can be afforded (e.g., revising a text without access to the feedback).

Given that metalinguistic codes are considered useful strategies for providing guided learning or fostering self-discovery skills (see Bitchener & Ferris, 2012; Lalande, 1982; van Beuningen et al., 2012), they have had a privileged position in L2 learning academic settings (see Leki, 1991). Conversely, while common in research practice, direct corrections may not be so pedagogically popular—let alone in L2 instructional settings where text revision is an important component. However, Bonilla et al. (2017, 2018, 2021) demonstrate that direct corrections can be effective L2 learning tools when they turn into input-providing and output-pushing strategies as a result of what learners are required to do with the feedback (e.g., by studying the direct corrections and then revising a text without access to them).

Pedagogical suggestion: If text revision after having studied the feedback does not involve passive copying, do not discard direct corrections as a viable pedagogical option in the L2 (writing) class.

Accuracy

Conclusion 7: Learners' grammatical accuracy may not suffer when they also attend to non-grammatical accuracy (or vice versa).

The findings in Bonilla et al. (2017, 2018, 2021) do not lend support to the alleged attentional issues that comprehensive CF poses (cf. Sheen et al., 2009). Overall, learners showed that their attention did not seem to be diverted when attending to issues of two different types (e.g., grammatical and non-grammatical) and that processing more or less comprehensive scopes was manageable for them as evidenced in their (short- and long-term) accuracy gains.

Pedagogical suggestion: Be positive about learners having enough attentional resources to attend to both grammatical and non-grammatical problems.

Conclusion 8: Although feedback on non-grammatical inaccuracies does not deteriorate learners' grammatical accuracy, the latter could be further maximized if learners' attention is devoted solely to grammar.

Bonilla et al. (2018, 2021) showed that irrespective of the comprehensive CF form (i.e., more or less broadened), grammatical accuracy was still achieved. Nevertheless, more accuracy gains during text revision and in new writings took place when the feedback emphasis was exclusively on grammatical problems.

Pedagogical suggestion: When the aim is enhancing grammatical accuracy, whenever possible allow exclusive attention to grammatical issues.

Conclusion 9: Separate error categories are not equally responsive to written CF.

If written CF worked in the same way irrespective of the targeted features, the patterns of response in Bonilla et al. (2021) would not have been as varied as they were. In fact, it was noted that the potentially influential factors in the results comprised more than the treatable/untreatable characteristic recurrently referred to in the literature (e.g., Bitchener & Ferris, 2012; Diab, 2015; Ferris, 1999). Instead, Bonilla et al. (2021) reveals an intricate amenability scenario that could have been due to as many variables as the amount of feedback, the number of targeted structures, the type of error, the complexity of an error, the feedback strategy, the knowledge that is tapped both in the feedback treatment, and the knowledge that is triggered in the revision instructions.

Pedagogical suggestion: Be realistic with the expected accuracy outcome and keep in mind that the short- and long-term accuracy of separate error categories will likely depend on something more than a teacher's choice of feedback strategy.

Cognitive Load

Conclusion 10: Rather than comprehensive CF alone, it is the combination of a broad scope in the feedback treatment and a low degree of explicitness in the feedback strategy that may cause cognitive overload.

It has been repeatedly suggested that the strain of asking learners to correct a wide number of errors could be too much for them to handle, hence, the claims about cognitive overload (see Bitchener, 2008; Ellis et al., 2008; Evans et al., 2010; Sheen, 2007). Interestingly, no significant differences between groups with the same feedback strategy but differing feedback scope were found, implying that there may be more to cognitive overload claims than just a comprehensive approach in and of itself. As a matter of fact, differences were found between direct CF groups and their metalinguistic counterparts and between a less comprehensive direct CF (focusing exclusively on grammatical issues) and a broader metalinguistic one (focusing on both grammatical and non-grammatical issues).

Pedagogical suggestion: To reduce L2 learners' chances of feeling overburdened, aim for a comprehensive CF form that targets issues of one type only (e.g., grammatical) and that treats those issues with a highly explicit feedback strategy (e.g., direct corrections).

Learner Factors

Conclusion 11: How learners react to a given written CF treatment and what feedback preferences they have could be influenced by their proficiency level.

Bonilla et al. (2017) found an association between learners' proficiency level and their feedback preferences, which indicates that high proficient learners may be more prone than low proficient ones to favor written CF with metalinguistic codes. Also, a relation between their proficiency level and their reactions to written CF exists.

The significant difference dealt with learners' attitudes towards the appropriateness of the feedback, which low proficient level learners felt more appropriate than high proficient ones.

Pedagogical suggestion: Avoid one-size-fits-all approaches to error correction and whenever possible factor in learners' proficiency level and preferences in feedback-related decisions.

Conclusion 12: Learners may have a more difficult time understanding codes than direct corrections.

Bonilla et al. (2018) provides evidence of the comprehensibility issues that metalinguistic codes could pose, where the two metalinguistic CF groups reported significantly lower ratings than both direct CF groups. The qualitative data suggest that part of the issue for the participants was not knowing how to go about error correction. In this sense, the results are a call for a reflection on what descriptive studies on error correction views have previously stated (e.g., Lee, 1997, 2005): that L2 teachers may have been overestimating their learners' ability to understand codes.

Pedagogical suggestion: Allot a training period to make sure learners are familiarized enough with the coding system. Also, given that learners' preferences for metalinguistic codes may be associated with their proficiency level, ensure that they feel/are linguistically ready to deal with the lower degree of explicitness of such a feedback strategy.

Conclusion

When Truscott (1996) initially asserted that correcting learners' grammatical errors was ineffective, harmful, and a waste of time, he based himself on the state of the field as he pointed out in his response to Ferris (1999). In the same controversial article, Truscott (1996) also conceded that future research could weaken his case. The author wrote then, "Future research on learner variables might show that certain subgroups of learners can benefit from correction under certain circumstances" (p. 361). In this respect, Truscott was right. Notwithstanding their limitations, Bonilla et al. (2017, 2018, 2021) as well as other sound empirical efforts (e.g., Diab, 2015; van Beuningen et al., 2012) have irrefutably rebutted the alleged futility of written CF. That is why today written CF can be construed as a useful pedagogical intervention that may contribute to maximizing learners' self-editing and revision skills and assisting L2 development under certain conditions. Finally, it is hoped that the present analysis serves as a springboard for thorough discussions that not only address the needs of feedback researchers but also answer the pedagogical queries of those at the forefront of the L2 writing classrooms: teachers.

References

- Benson, S., & DeKeyser, R. (2018). Effects of written corrective feedback and language aptitude on verb tense accuracy. *Language Teaching Research*, 00(0), 1–25. <https://doi.org/DOI:10.1177/1362168818770921>
- 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. (2012a). A reflection on ‘the language learning potential’ of written CF. *Journal of Second Language Writing*, 21(4), 348–363. <https://doi.org/10.1016/j.jslw.2012.09.006>
- Bitchener, J. (2012b). Written Corrective Feedback for L2 Development: Current Knowledge and Future Research. *TESOL Quarterly*, 46(4), 855–860. <https://doi.org/10.1002/tesq.62>
- Bitchener, J. & Ferris, D. (2012). *Written corrective feedback in second language acquisition and writing*. Routledge.
- 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. (2010a). 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. (2010b). 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., & Storch, N. (2016). *Written Corrective Feedback for L2 Development* (Vol. 96). Multilingual Matters.
- 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. <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. <https://doi.org/10.1111/lang.12295>
- Bonilla, M., Van Steendam, E., Speelman, D., & Buyse, D. (2021). Comprehensive corrective feedback in second language writing: The response of individual error categories. *Journal of Writing Research*, 13(1), 31–70. <https://doi.org/10.17239/jowr-2021.13.01.02>
- Bruton, A. (2010). Another reply to Truscott on error correction: Improved situated designs over statistics. *System*, 38(3), 491–498. <https://doi.org/10.1016/j.system.2010.07.001>
- Byrnes, H., & Manchón, R. M. (2014). Task-based language learning. In *Task-based language learning: Insights from and for L2 writing*. John Benjamins Publishing Company.
- Diab, N. (2015). Effectiveness of written corrective feedback: Does type

- of error and type of correction matter? *Assessing Writing*, 24, 16–34. <https://doi.org/10.1016/j.asw.2015.02.001>
- Ellis, R. (2009a). Corrective feedback and teacher development. *L2 Journal*, 1(1), 3–18.
- Ellis, R. (2009b). A typology of written corrective feedback types. *ELT Journal*, 63(2), 97–107. <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. <https://doi.org/10.1017/S0272263109990544>
- Ellis, R., & Shintani. (2014). *Exploring language pedagogy through second language acquisition research*. Routledge.
- Ellis, 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>
- Enginarlar, H. (1993). Student response to teacher feedback in EFL writing. *System*, 21(2), 193–204.
- 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>
- Ferris, D. (1995). Student reactions to teacher response in multiple-draft composition classrooms. *TESOL Quarterly*, 29(1), 33. <https://doi.org/10.2307/3587804>
- 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.
- 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. <https://doi.org/10.1016/j.jslw.2004.04.005>
- Ferris, D. (2010). Second language writing research and written corrective feedback in SLA. *Studies in Second Language Acquisition*, 32(02), 181–201. <https://doi.org/10.1017/S0272263109990490>
- 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
- Guénette, D. (2007). Is feedback pedagogically correct? *Journal of Second Language Writing*, 16(1), 40–53. <https://doi.org/10.1016/j.jslw.2007.01.001>
- 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>
- Hedgcock, J., & Lefkowitz, N. (1994). Feedback on feedback: Assessing learner receptivity to teacher response in L2 composing. *Journal of Second Language Writing*, 3(2), 141–163.
- Hedgcock, J., & Lefkowitz, N. (1996). Some input on input: Two analyses of student response to expert feedback in L2 writing. *The Modern Language Journal*, 80(3), 207–308.

- Hyland, F. (2001). Providing effective support: Investigating feedback to distance language learners. *Open Learning*, 16(3), 233–247.
- Hyland, K., & Hyland, F. (2006). Feedback on second language students' writing. *Language Teaching*, 39(02), 83. <https://doi.org/10.1017/S0261444806003399>
- Karim, K., & Nassaji, H. (2020). The revision and transfer effects of direct and indirect comprehensive corrective feedback on ESL students' writing. *Language Teaching Research*, 24(4), 519–539. <https://doi.org/10.1177/1362168818802469>
- Kurzer, K. (2022). Accuracy Gains from Unfocused Feedback: Dynamic Written Corrective Feedback as Meaningful Pedagogy. *Journal of Language and Education*, 8(4), 102–116. <https://doi.org/10.17323/jle.2022.13380>
- Lalande, J. F. (1982). Reducing composition errors: An experiment. *The Modern Language Journal*, 66(2), 140–149
- Lee, I. (1997). ESL learners' performance in error correction in writing: Some implications for teaching. *System*, 25(4), 465–477.
- Lee, I. (2005). Error correction in the L2 writing classroom: What do students think? *TESL Canada Journal*, 22(2), 1–16.
- Lee, I. (2008). Student reactions to teacher feedback in two Hong Kong secondary classrooms. *Journal of Second Language Writing*, 17(3), 144–164. <https://doi.org/10.1016/j.jslw.2007.12.001>
- Leki, I. (1991). The preferences of ESL students for error correction in college level writing classes. *Foreign Language Annals*, 24(3), 203–218.
- Lyster, R. (2004). Differential effects of prompts and recasts in form-focused instruction. *Studies in Second Language Acquisition*, 26(03). <https://doi.org/10.1017/S0272263104263021>
- McMartin-Miller, C. (2014). How much feedback is enough?: Instructor practices and student attitudes toward error treatment in second language writing. *Assessing Writing*, 19, 24–35. <https://doi.org/10.1016/j.asw.2013.11.003>
- Nassaji, H. (2007). Elicitation and reformulation and their relationship with learner repair in dyadic interaction. *Language Learning*, 57(4), 511–548. <https://doi.org/10.1111/j.1467-9922.2007.00427.x>
- Oladejo, J. A. (1993). Error Correction in ESL: Learner's Preferences. *TESL Canada Journal*, 10(2), 71–89.
- 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>
- Robinson, P. (2005). Cognitive complexity and task sequencing: Studies in a componential framework for second language task design. *IRAL*, 43, 1–32.
- Sampson, A. (2012). “Coded and uncoded error feedback: Effects on error frequencies in adult Colombian EFL learners' writing”. *System*, 40(4), 494–504. <https://doi.org/10.1016/j.system.2012.10.001>
- Schmidt, R. W. (1990). The role of consciousness in second language learning. *Applied Linguistics*, 11(2), 129–158.

- Schmidt, R. W. (2001). Attention. En P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3–32). Cambridge University Press.
- Sheen, Y. (2004). Corrective feedback and learner uptake in communicative classrooms across instructional settings. *Language teaching research*, 8(3), 263–300.
- 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(02), 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>
- Skehan, P. (1998). *A Cognitive Approach to Language Learning*. Oxford University Press.
- Skehan, P. (2003). Task-based instruction. *Language Teaching*, 36(1), 1–14. <https://doi.org/10.1017/S026144480200188X>
- 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. <https://doi.org/10.1017/S0272263109990532>
- Swain, M., & Lapkin, S. (2002). Talking it through: Two French immersion learners' response to reformulation. *International Journal of Educational Research*, 37(3–4), 285–304. [https://doi.org/10.1016/S0883-0355\(03\)00006-5](https://doi.org/10.1016/S0883-0355(03)00006-5)
- Swain M. (2006). Linguaging, Agency and Collaboration in Advanced Second Language Proficiency. In Byrnes H. (Ed.), *Advanced Language Learning: The Contribution of Halliday and Vygotsky*, 95-108. Continuum
- Truscott, J. (1996). The case against grammar correction in L2 writing classes. *Language learning*, 46(2), 327–369.
- Truscott, J. (1999). The case for “The case against grammar correction in L2 writing classes”: A response to Ferris. *Journal of Second Language Writing*, 8(2), 111–122.
- Truscott, J. (2001). Selecting errors for selective error correction. *Concentric: Studies in English Literature and Linguistics*, 27(2), 93–108.
- 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. (2010). Corrective feedback in L2 writing: Theoretical perspectives, empirical insights, and future directions. *International Journal of Educational Research*, 37(3–4), 285–304. [https://doi.org/10.1016/S0883-0355\(03\)00006-5](https://doi.org/10.1016/S0883-0355(03)00006-5)

- tional Journal of English Studies*, 10(2), 1–27.
- van Beuningen, C., De Jong, N. H., & Kuiken, F. (2008). The effect of direct and indirect corrective feedback on L2 learners' written accuracy. *ITL-International Journal of Applied Linguistics*, 156, 279–296.
- van Beuningen, C., De Jong, N. H., & 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. <https://doi.org/10.1111/j.1467-9922.2011.00674.x>