Measuring the Impact of Written Corrective Feedback (WCF): The Methodological ins and outs

Benjamin Carcamo[1]

[1] Facultad de Comunicaciones y Artes, Universidad de Las Américas, Chile

Correspondence: Benjamin Carcamo, Manuel Montt 948, 7500975 Providencia, Santiago, Chile. Orcid: https://orcid.org/0000-0001-7330-9007. E-mail: bcarcamo @udla.cl

Received: February 20, 2024 Accepted: May 23, 2024 Online Published: July 30, 2024

doi:10.5430/wjel.v14n6p270 URL: https://doi.org/10.5430/wjel.v14n6p270

Abstract

The interest in the study of written corrective feedback (WCF) has led to the identification of two controversial areas: the impact of the different types of WCF and the methods used in the studies. The present study aims to contribute to the clarification of the latter by examining the methods section of recently published research articles. Its objective is to review the current state of the experimental research designs in WCF studies. For this, eleven published studies were randomly selected and analyzed considering as inclusion criteria that they were empirical studies published in journals indexed in Web of Science and/or Scopus during the last eight years. The findings have been grouped into two categories: research designs and data analysis procedures. Regarding research design, there is a growing concern for ecological validity, the inclusion of delayed post-tests, and the implementation of training on feedback types for students. Concerning data analysis, the choice of a specific formula for error quantification was identified as a crucial decision researchers interested in WCF must make. It is hoped that this study will guide future research, ensuring a solid methodological design in line with current publications in prestigious journals. By following the trends identified in recent prestigious research, it is more likely that research can address the concerns surrounding the effectiveness of different types of WCF.

Keywords: error quantification, experimental studies, methodology, second language acquisition, written corrective feedback

1. Introduction

The study of written corrective feedback (WCF) has been steadily growing in the last decade**.** A recent bibliometric analysis by Crosthwaite et al. (2022) examined 493 L2 WCF-related articles with the aim of mapping trends in the field from a longitudinal perspective. Their findings indicate a rising interest in teacher and peer feedback, stakeholders' perceptions, automated writing evaluation, the differences between direct and indirect feedback, and learner engagement. Additionally, research from the USA was found to be the most frequently published. However, there is a noticeable increase in studies originating from other countries such as China and Iran. Despite this, WCF research has also been presented as a controversial topic in the investigation of second language acquisition, primarily due to its questionable effectiveness (Fazio, 2001; Truscott, 1996; Truscott & Hsu, 2008) and its lack of methodological thoroughness (Liu & Brown, 2015; Storch, 2010). Regarding the first concern, numerous studies have sought to elucidate the impact of different types of WCF in written production (Diab, 2015; Ferreira, 2017; Ghazi & Zamanian, 2016; Liu, 2008; Tang & Liu, 2018).

In terms of methodological development in WCF studies, Storch (2010) noted significant progress with respect to the studies done at the end of the 20th century. The early studies on WCF exhibited weaknesses such as absence of a control group, exclusive reliance on drafts without the production of new writing, utilization of tasks that did not require attention to linguistic accuracy (e.g., use of reflection journals) and a failure to focus on 'treatable' errors. In addition, these studies lacked comparability due to the vague descriptions of participants, the excessive variety of treatments, and the diverse methods for measuring linguistic accuracy. Recent studies have addressed these shortcomings; however, this greater experimental refinement has introduced other types of challenges that diminish the ecological validity of the research. Some of these limitations include sporadic interventions (one-off treatments), insufficient attention to the goals of the learners, and ignoring the attitudes of those involved towards WCF. In this context, Riazantseva (2012) suggests that although some deficiencies of early studies have been dealt with, there is a risk of losing the strengths of this type of research.

Liu and Brown (2015) examined 32 published studies and 12 theses to account for the methodological thoroughness of studies in WCF. Among their findings, two stand out: the little detail with which investigators report the research context and the lack of consistency in the methodology section. Similar to Storch (2010), Liu and Brown (2015) highlighted the low ecological validity in the studies and a variety of ways to quantify the improvement in linguistic accuracy. Furthermore, the authors observed a lack of control over the WCF type variable, as investigators often considered combinations of different types as part of the treatment for the same group.

In this context, it is evident that the methodological procedures followed by researchers interested in WCF must be further examined and systematized. The nuanced distinctions between early and contemporary studies underscore the necessity for further methodological refinement, which can be achieved by analyzing what is currently being published in the field. Otherwise, there is an imminent risk of

losing sight of the advancements recognized by the scholarly community as appropriate methodological choices.

Given this challenging situation, it becomes imperative to establish a clear understanding on how the impact of WCF is currently being researched. This clarity is essential for researchers to make well-informed decisions (Carcamo, 2020; Ellis, 2009). Consequently, the objective of this article is to identify the conventional methods in which researchers have studied and reported the effects of WCF. Specifically, this work focuses on research published over a span of eight years (2015-2023) in journals indexed in Web of Science (WoS) as to give an account of the state of the art regarding the methodology employed in studies on WCF. The present study will contribute to the field by identifying clear trends that can guide new researchers in establishing a robust methodological framework for conducting empirical research on WCF effectiveness.

2. Method

2.1 Corpus

Specific criteria were established for the selection of the articles that would be analyzed. Firstly, a timeframe of eight years spanning from 2015 to 2023 was set for the inclusion of articles in the analysis, and *written corrective feedback* was utilized as the key term for article search purposes. Secondly, the prestige of the journals where the research was published was used as a criterion for the collection of articles. Journal indexing was regarded as an indicator of adherence to the standards of the scientific community; therefore, articles that had been published in journals indexed in Web of Science (WoS) and/or Scopus were chosen. Thirdly, it was determined that the articles had to be of an empirical nature and conducted with human participants. Lastly, the study had to be conducted in an EFL context. Based on the previous criteria, a random sample of eleven research articles was chosen (Appendix A). The procedure was the following: We assigned a number to each of the articles from the sampling pool, then, the website randomizer.org was used. The final selection of eleven articles is similar to what has been observed in other systematic literature reviews with the intention of exploring in depth the content of the articles compiled (Gray, 2021).

2.2 Data Analysis

The analysis performed followed two criteria: Research design and data analysis. The criterion of research design was understood as the procedures used to obtain the information needed to solve the research problem (Hernández-Sampieri & Mendoza Torres, 2018). In other words, the way in which the researcher organizes the type of study, the data collection procedures, the phases of the research process and the approach from which the research is proposed. The second category that guided the analysis was the processing and analysis of the data, which considers the way in which the errors of the students are quantified, and the statistical tests used in the research. To systematize the analysis, we combined a deductive with an inductive approach to the corpus. Deductive in the sense that we analyzed the corpus considering the categories of research design and data analysis, but inductive in terms of identifying the features present in each of the two categories.

It should be noted that other studies are occasionally used to support the importance and validity of some of the statements made regarding research in WCF in selected publications. Thus, even though eleven studies form the core of the analysis, others are cited and referenced to enhance the robustness of the claims made in terms of the implications the findings have for the field.

3. Results and Discussion

3.1 What Are the Research Designs Used?

As highlighted by Storch (2010), one of the main challenges in research on WCF is ensuring ecological validity. Four trends have been identified to address this concern in the analyzed studies. Firstly, the implementation of quasi-experimental designs. Secondly, the inclusion of multiple samples by usually conducting a pre-test, a post-test, and a delayed post-test. Third, the use of training on WCF for students to enhance their understanding of what they will receive during the intervention part of the research. Fourthly, the manner in which writing is elicited from the students. While there is agreement on the practices that researchers tend to follow to address the first two challenges, consensus is not as clear regarding the latter two.

One of the experimental designs that helps preserve the ecological validity of a study during the intervention is the quasi-experimental design. These studies are characterized by the manipulation of independent variables while maintaining the integrity of the groups of subjects used for the study (Hernández-Sampieri & Mendoza Torres, 2018). Among the articles analyzed, 81.8% follow this design. Specifically, researchers often keep one group of students intact while exposing one or two others to a particular intervention as the experimental group. This approach allows researchers to compare the impact of the application of different types of WCF among different courses without the students feeling part of a study that interferes with the normality of their classes. The majority of the studies tend to choose two experimental groups and one control group.

Diab (2015), for example, selected three university courses of 20, 19, and 18 subjects. Two of these groups received indirect WCF, one with metalinguistic comments and the other without them. The third group served as a control and did not receive WCF. Similarly, Frear and Chiu (2015) also made use of 3 courses which they classified as 2 experimental groups and one of control. The experimental groups of both studies received indirect WCF, one of them focused and the other unfocused, while in the control group the students had to review their own writings (*self-feedback*). Likewise, Shintani and Ellis (2015) selected four English courses from a Japanese university, keeping them intact while randomly assigning the treatment each participant would receive as WCF. Other studies in which this approach can be observed are those by Zhang (2017), Guo and Barrot (2019), Bagheri and Rassaei (2022), Supiani et al. (2023) and Stefanou and Révész (2015).

Regarding the procedure to study the effect of WCF, researchers tend to follow a three-stage evaluation process. A pre-test, an immediate post-test, and a delayed post-test. During the pre-test, students are tasked with writing a text of a specific genre, for which they receive initial WCF. Measuring the errors made in this first piece of writing helps the researcher determine the baseline of students' proficiency before the intervention. In the immediate post-test, students produce a new representative text for the same task after receiving WCF during the intervention. Finally, in the delayed post-test, students generate a final text weeks after performing the immediate post-test, so that researchers can determine if their learning lasted beyond the intervention. This three-stage procedure is commonly employed in WCF studies (Bagheri and Rassaei, 2022; Bitchener, 2008; Bitchener & Knoch, 2009; Diab, 2015; Frear & Chiu, 2015; Guo & Barrot, 2019; Stefanou & Révész, 2015; Zhang, 2017). It is important to note that the delayed post-test is typically conducted around twenty days after the application of the immediate post-test to verify whether learning was permanent or not, which is a common focus in WCF studies. In the case of Diab's study (2015), the time between the immediate post-test and the delayed was longer, reaching nine weeks, so this alternative timeframe should also be taken into account as a possible standard by researchers.

What students do when they receive the pre-test with feedback is also a crucial aspect to consider during the design of the study. In this regard, two paths can be identified in line with the responses to the WCF proposed by Ellis (2009). On the one hand, students can be asked to correct the indicated errors (Diab, 2015) and, on the other hand, to study them for a limited time that can range from five (Frear & Chiu, 2015; Guo & Barrot, 2019; Stefanou & Révész, 2015) to twenty minutes (Zhang, 2017). It is worth noting that only Stefanou and Révész (2015) indicate how the review time was defined, referring to the results of a pilot study. In a similar vein, Shintani and Ellis (2015) included in their study whether students reviewed the WCF they received as a variable, demonstrating that performing this action affected the quality of the writing they did afterwards.

The role of training in WCF studies is addressed inconsistently in the analyzed studies. Diab (2015) points out that his intervention used two sessions to address the categories of error that would be corrected and their symbology (Pr. Agr = Concordance pronoun and W.W. = Wrong words). Students were informed about nuances, such as how receiving W.W. could encompass a vocabulary error related to contextual connotations, word order issues, the degree of formality, the redundancy caused by the word or the nonexistence of the used word. This type of training in the case of Bagheri and Rassaei (2022) consisted of three sessions which were held during a week. On the other hand, such as Frear and Chiu (2015), Shintani and Ellis (2015), and Supani et al. (2023) do not specify any training session for students to better understand the WCF that will be given to them.

Another point to consider in the research design is the selection of tasks that students must perform. Two trends can be identified: The first trend involves using tasks that naturally occur in the class and a second one that prioritizes creating tasks that increase the likelihood of specific errors. From the first approach, Park et al. (2016), for instance, used the students' diaries, which were updated week by week, for analysis. Similarly, Kim et al. (2020) made use of three writing genres that were relevant to the lessons covered in class, designing tasks as information gaps to promote communication that resembled real-world activities where these genres were employed. On the other hand, other researchers focus on designing activities that seek to promote the use of specific forms in tasks designed for the study (Bagheri & Rassaei, 2022; Frear & Chiu, 2015; Guo & Barrot, 2019; Shintani & Ellis, 2015; Stefanou & Révész, 2015). Frear and Chiu (2015), for example, implemented an activity that sought to simulate the dynamics of a regular class. Students first read a news story focusing on the unfamiliar vocabulary. Then, the students gathered in groups to solve the difficulties with the vocabulary, and, if they still had problems, they asked the teacher for help. After this process, students read the text again. Finally, the students received a series of images that represented the content of the news along with the first sentence of the text they had to create to retell the story. This sentence was formulated in the past to elicit students to use the forms that researchers sought to study.

The identified research trends show that researchers have made progress towards addressing methodological concerns such as ecological validity and the impact of WCF. The prevalence of quasi-experimental designs and the incorporation of training sessions for students indicate concrete efforts to enhance the validity of research methodologies. This emerging trend holds significance in enhancing the comparability of studies and the distinction between the effects of various types of feedback. Without adequate training, both teachers and students might misinterpret types of feedback, potentially altering their use within the experimental studies in which they partake. On the other hand, a lingering area of ambiguity pertains to the criteria for selecting appropriate writing tasks for data collection. Prior research (see Bao, 2015; Brown et al., 2023) has shown that different tasks demand different language sub-skills. Therefore, careful consideration of task demands is imperative to elucidate the factors that contribute to the effectiveness of specific types of WCF.

How is the data analyzed?

To assess the improvement in linguistic accuracy, researchers commonly follow two main paths: using tasks evaluated with a rubric that focuses on overall writing quality rather than specific linguistic errors, or the objective quantification of specific errors. An example of this approach can be observed in Stefanou and Révész's (2015) study where two tasks were designed to evaluate the use of articles with specific and generic referents. The researchers used a text summary test and a truth value judgment test to assess students' writing. In the text summary task, participants read a short text and then wrote the description of some of the animals mentioned in it according to some photographs that were given to them. The truth-value judgment task, a test of receptive knowledge, required students to determine whether a statement about the text was true or not. Alternatively, also from the perspective of an unfocused approach to correction, Zhang (2017) employed an analytical rubric created by Kobayashi and Rinnert (1992) to assess the quality of the writings and the possible improvement during the intervention.

On the other hand, the option of quantifying the linguistic accuracy of writings involves beginning with error coding. Before quantifying students' errors in their writings, researchers calculate the level of agreement between annotators using statistics such as the Pearson Product-Moment Correlation Coefficient. Across the studies analyzed, the reported correlation coefficients ranged from r = .79 (Diab, 2015) to r = .99 (Frear & Chiu, 2015). After this stage, conventional formulas are applied. One commonly used formula is the Target-Like Use Analysis (Bagheri and Rassaei, 2022; Endley & Karim, 2022; Frear & Chiu, 2015; Shintani and Ellis, 2015), which originates from Pica's morpheme error correction proposal (1983). The Target-Like Use Analysis Formula involves dividing the number of times the structure was used correctly by the times in which it had to be used compulsorily plus instances of use in non-mandatory contexts. This result is then multiplied by a hundred in order to obtain a percentage. Figure 1 illustrates the described calculation.

$$\frac{\text{Number of accurate instances of grammar use}}{\text{Number of obligatory contexts + number of overused uses}} \text{ X } 100$$

Figure 1. Target-Like Use Analysis Formula

Clearly this formula is useful in the case of applying WCF to focused errors. Bagheri and Rassaei (2022), for instance, calculate learners' article usage (the grammar points in which their study focused). The calculation involved dividing the number of correct uses by the obligatory context plus the additional uses of articles, then multiplying the result by 100. On the other hand, when calculating the total percentage of linguistic accuracy, which may not be convenient to do in each linguistic instance due to the multiple possibilities that language provides, a more comprehensive formula is required. For the calculation of the total linguistic accuracy, Frear and Chiu (2015), for example, propose the formula shown in Figure 2, which was first provided by Ellis and Barkhuizen (2005).

$$\frac{\text{Number of error-free clauses}}{\text{Number of total clauses}} \text{ X } 100$$

Figure 2. Target-Like Use Analysis Formula

As shown in Figure 2, this formula implies the numbering of the clauses of the text written by the student in order to identify those that do not present errors. According to Frear and Chiu (2015), an advantage of this formula is that its result is a percentage. Similarly, other formulas for error ratio have been used to estimate the overall accuracy. Van Beuningen, et al. (2012) used the following formula:

$$\frac{\text{Number of linguistic errors}}{\text{Total number of words}} \text{ X } 10$$

Figure 3. Overall accuracy formula

The use of a 10-word ratio instead of a 100-word ratio was chosen due to the limited length of the text produced in the researchers' study. Consequently, the multiplier of the formula can be said to be 10 when the task involves short pieces of writing and 100 when the writing produced is long.

Once errors have been quantified, the typical procedure in the examined studies is to carry out statistical analyses to confirm the effect of WCF on students' writing. The choice of statistical analysis depends on the design that has been proposed. Given that a common practice is working with two experimental groups (each receiving different WCF) and a control group, one of the most frequently used statistical analyses is one or more types of the Analysis of Variances Test (ANOVA). The one-way ANOVA test is often used to examine the means of groups, such as in the pre-test, to establish whether the groups can be considered equal. Therefore, its application is relevant for the analysis of between-subject variables. Moreover, this statistical test can be applied to note if these groups that were initially similar demonstrate differences after the intervention in the immediate or delayed post-tests, which can be attributed to the WCF provided. In one of the studies (Guo & Barrot, 2019), an ANCOVA test was used, which works in the same way an ANOVA test does while also allowing for the consideration of a covariate in the analysis.

Furthermore, most of the investigations examined within-subject variation through repeated-measures ANOVA models, often involving one or two factors. The term 'repeated measures' is used for studies in which the same entities participate in all the conditions of an experiment or provide data at multiple time points (Field, et al., 2012). In the context of WCF studies, repeated measures ANOVA typically fall into the

latter category. These studies commonly incorporate three or more groups whose means and variances are to be contrasted over time. That is, there is an independent grouping variable of more than two levels (usually type of feedback) and groups whose linguistic accuracy is compared at three different times.

After quantifying errors and conducting statistical tests, it's not only possible to determine if there are statistically significant differences in the intervention but also to gauge the magnitude of these differences. The measure corresponding to the magnitude of a difference is known as the effect size (Cohen, 1988). Effect sizes are relevant since they indicate the impact of a particular independent variable on a dependent variable. Moreover, they can be utilized to estimate sample size using specialized software, such as G*Power (Faul et al., 2009), which, in turn, helps safeguard the statistical power of the study (Lenth, 2007). Table 1 displays some of the effect size values from the studies analyzed, which have been segmented depending on the way in which they approached writing improvement.

Table 2. Effect sizes (η2) in WCF studies

Study	Variables	Partial η2
Diab (2015)	Accuracy in pronoun agreement / Within-subject	
	a) Test sessions	a) 0.210
	b) Test sessions x treatment	b) 0.155
Diab (2015)	Lexical errors / Within-subject	
	a) Test sessions	a) 0.55
	b) Test sessions x treatment	b) 0.148
Frear & Chiu (2015)	a) Indirect focused WCF on regular past	a) 0.51
	b) Indirect unfocused WCF on regular past	b) 0.53
Frear & Chiu (2015)	a) Indirect focused WCF on irregular past	a) 0.47
	b) Unfocused indirect WCF on irregular past	b) Not statistically significant
Stefanou & Révész (2015)	a) Time x group in text summary test	a) 0.22
	b) Time x group in truth value judgment test	b) 0.07
Zhang (2017)	a) Group (score in rubric)	a) 0.069
	b) Time (score in rubric)	b) 0.465
	c) Time x group (score in rubric)	c) 0.126
Zhang (2017)	a) Group (score in content criterion)	a) 0.089
	b) Time (score in content criterion)	b) 0.415
	c) Time x group (score in content criterion)	c) 0.114
Zhang (2017)	a) Group (score in organization criterion)	a) Not statistically significant
	b) Time (score in organization criterion)	b) 0.404
	c) Time x group (score in organization criterion)	c) 0.121
Zhang (2017)	a) Group (score in language criterion)	a) 0.117
	b) Time (score in language criterion)	b) 0.366
	c) Time x group (score in language criterion)	c) 0.071
Guo & Barrot (2019)	a) Regular past tense	a) Not statistically significant
	b) Prepositions indicating space	b) Not statistically significant
Kim et al. (2020)	a) Scores in post-test (in favor of SWCF groups)	a) 0.13
	b) Indirect SWCF vs Direct SWCF	b) Not statistically significant
Bagheri & Rassaei (2022)	a) Time	a) 0.69
	b) Feedback conditions	b) 0.50
	c) Time x written CF	c) 0.42
Endley & Karim (2022)	a) Accuracy rates in Timed Grammaticality Judgment Tasks (Articles only)	a) Not statistically significant
	b) Accuracy rates in Timed Grammaticality Judgment Tasks (Prepositions only)	b) Not statistically significant
	c) Accuracy rate in writing task (articles)	c) Not statistically significant
	d) Accuracy rate in writing task (prepositions)	d) Not statistically significant
Supiani et al. (2023)	a) Direct WCF and control group	a) 0.3
	b) Metalinguistic Explanation and control group	b) 0.2
	c) Direct WCF and Metalinguistic Explanation	c) Not statistically significant

As observed in the table, most studies have shown that WCF has a statistically significant effect on students' writing and grammar accuracy. According to Pallant (2016), a small effect size for ANOVA tests is represented by an Eta squared (η2) of .01, a medium effect size by .06 and a large one by .138. Notably, it can be stated that most of the studies have proved the importance of using WCF as the reported effect sizes were predominantly large and medium. It is worth mentioning that none of these studies shared a small effect size once the result was statistically significant.

In the case of comparisons between 2 types of students, the T Test of Independent Samples was appropriately chosen. The size of the effect in this case is the Cohen's *d*. Table 2 presents the reported effect sizes.

Table 3. Effect size Cohen's *d* in WCF studies

Study	Variables	Cohen's *d*
Park, Song, & Shin (2016)	a) Distribution of error types based on L2 level (elementary vs. intermediate).	a) Not significant.
	b) Distribution of error types based on heritage language status	b1) 1.76 = Orthographic errors b2) 1.26 = Tense and conjugation errors b3) 1.24 = Particle errors b4) 0.77 = Lexical errors
Park, Song, & Shin (2016)	a) Perception of errors based on L2 level (elementary vs. intermediate)	a) Not significant.
	b) Perception of errors based on heritage language status	b1) 1.44 = Orthographic errors b2) 2.02 = Particle errors b3) Not significant for tense, conjugation or lexical.
Park, Song, & Shin (2016)	a) Self-correction of errors based on L2 level (elementary vs. intermediate)	a1) 0.91 = Particle errors a2) Not significant for orthographic, tense/conjugation or lexical.
	b) Self-correction of errors based on heritage language status	b) Not significant
Bagheri & Rassaei (2022)	a) Direct vs Indirect CF b) Direct CF vs Control c) Indirect CF vs Control	a) 0.76 b) 2.6 c) 2.1

When observing the table, it is noticeable that a significant majority of the studies report statistically significant findings. According to Pallant (2016), the interpretation of Cohen's *d* is as follows: the effect size is small when the value is 0.2, medium if it is 0.5, and large if it is 0.8. Most studies shown in the preceding table obtained large and medium effect size. This suggests a substantial impact of WCF on students' writing abilities, which underscores the importance of WCF in EFL classrooms and second language acquisition research. These findings carry significant implications for EFL teaching and learning, as they suggest that the implementation of WCF strategies can yield benefits in enhancing students' writing proficiency. EFL educators can use these results to inform their instruction. However, despite the promising findings, gaps persist in our understanding of the optimal implementation of feedback, considering that studies tend to focus on specific grammatical aspects overlooking the role that the task selected for data collection can play. Namely, from an EFL teaching perspective, the importance of giving feedback on students' writing would go beyond specific mistakes related to language. Instead, the focus would be on better achieving the task itself (Bygate et al., 2015). It is possible that incorporating alternative perspectives on WCF, which broaden the understanding of WCF towards a comprehensive examination of the text (Carcamo, 2020) can shed light on the extent to which WCF can positively impact students' writing.

4. Conclusion

WCF studies have grown considerably in recent years due to their perceived utility in second language teaching. However, despite their increasing importance, challenges persist regarding the degree of effectiveness of this technique and the definition of the most efficient investigation standards and protocols. In this article the second of these challenges has been addressed through the analysis of prestigious experimental research in the field, unveiling trends commonly observed in WCF studies, which should be followed by researchers.

Among the findings the importance of quasi-experimental design and the incorporation of natural tasks to ensure ecological validity have been highlighted. Additionally, while sporadic treatments remain prevalent, the greater rigor in experimental designs has contributed to accumulating evidence supporting the efficacy of WCF interventions, evidenced by consistently reported medium and large effect sizes. Consequently, criticism, such as those voiced by Storch (2010), regarding the excessive use of this type of intervention appear to be losing relevance in contemporary discussions. Moreover, this research emphasizes key elements crucial for an effective design. These include the application of a sequence of evaluations involving a pre-test, immediate post-test and delayed post-test. The use of two or more experimental groups and a control group facilitates the comparison of the effectiveness of different types of WCF.

Two less standardized elements in WCF research have surfaced: The student training phase and the error quantification formulas used for error quantification. The training phase is omitted in most of the analyzed studies, which is curious considering that students need to understand the feedback they will receive effectively as to fully process it. With regard to error quantification formulas, it is surprising, given the limited scope of the selected studies, that there is variation in error quantification formulas employed. This finding aligns with the results of the meta-analysis by Liu and Brown (2015).

The present study makes two notable contributions to the field. Firstly, it offers significant insights into how to refine methodological procedures associated with WCF research. By enhancing the quality of the methods used in these studies, teachers and researchers can gain a clearer understanding of the most effective approaches to proving feedback on writing. Secondly, this research helps us move a step closer toward elucidating the ongoing controversies surrounding the impact of different WCF types and experimental methodologies. Our findings provide guidance for researchers about the positive practices currently being used in WCF research, which should ultimately enhance the validity and reliability of future studies, thereby helping address the controversies outlined in prior research.

Despite its contributions, this study has certain limitations. One of the most important limitations is the narrowed focus on aspects related to

research design and error quantification, which led us to make decisions that involved neglecting other important features of designing WCF research, such as the number of participants in the sample or the specific characteristics of the feedback given. Another limitation is the methodological decision of examining 11 studies and not considering taking a broader scope. While some systematic reviews adhere to similar numbers, future studies could explore the option of conducting meta-analyses encompassing a broader sample size to contribute with valuable data that can inform trends in the field as it has been done in meta-analyses such as the ones conducted by Crosthwaite et al. (2022), Liu & Brown (2015), and Storch (2010).

Future studies should not only focus on refining methodologies, but also explore new avenues for inquiry. It would be beneficial for the field to conduct studies investigating the long-term of WCF on language proficiency, along with the role of digital tools in this process. Furthermore, to gain further knowledge on the efficacy of WCF, we suggest researchers consider examining the impact of different types of WCF considering students' and teachers' perceptions and understanding of their impact in writing.

It is expected that the results of this work will help guide future research, offering key critical points that should be considered both in the design of WCF studies and in the analysis of the results obtained. These insights can help new as well as established researchers refine their methodologies, enhancing the quality of the research being conducted. A greater methodological awareness is one of the fundamental dimensions to solve the doubts that have been raised regarding the usefulness of WCF.

Acknowledgments

I would like to thank the IDI research group for their support.

Authors contributions

BC was responsible for study design, revising, data collection, and drafting the manuscript.

Funding

Not applicable

Competing interests

The author declares that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Informed consent

Obtained.

Ethics approval

The Publication Ethics Committee of the Sciedu Press.

The journal's policies adhere to the Core Practices established by the Committee on Publication Ethics (COPE).

Provenance and peer review

Not commissioned; externally double-blind peer reviewed.

Data availability statement

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

Data sharing statement

No additional data are available.

References

Bagheri, M., & Rassaei, E. (2022). The Effects of Two Forms of Written Corrective Feedback and Ambiguity Tolerance on EFL Learners' Writing Accuracy. *English Teaching & Learning*, *46*, 19-38. https://doi.org/10.1007/s42321-021-00082-6

Bao, G. (2015). Task type effects on English as a Foreign Language learners' acquisition of receptive and productive vocabulary knowledge. *System*, *53*, 84-95. https://doi.org/10.1016/j.system.2015.07.006

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

Bitchener, J., & Knoch, U. (2009). The contribution of written corrective feedback to language development: A ten month investigation.

Applied Linguistics, *31*(2), 193-214. https://doi.org/10.1093/applin/amp016

Brown, D., Liu, Q., & Norouzian, R. (2023). Effectiveness of written corrective feedback in developing L2 accuracy: A Bayesian meta-analysis. *Language Teaching Research*, *0*(0). https://doi.org/10.1177/13621688221147374

Bygate, M., Norris, J., & van den Branden, K. (2015). Task-Based Language Teaching. In C. A. Chapelle (Ed), *The Encyclopedia of Applied Linguistics* (pp. 1-8). New York: John Wiley & Sons, Ltd. https://doi.org/10.1002/9781405198431.WBEAL1467.

Carcamo, B. (2020). Classifying written corrective feedback for research and educational purposes: A typology proposal. *Profile*, *22*(2), 211-222. https://doi.org/10.15446/profile.v22n2.79924

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

Crosthwaite, P., Ningrum, S., & Lee, I. (2022). Research trends in L2 written corrective feedback: A bibliometric analysis of three decades of Scopus-indexed research on L2 WCF. *Journal of Second Language Writing*, *58*, 100934. https://doi.org/10.1016/j.jslw.2022.100934

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. (2009). A typology of written corrective feedback types. *ELT Journal*, *63*(2), 97-107. https://doi.org/10.1093/elt/ccn023

Ellis, R., & Barkhuizen, G. (2005). *Analyzing learner language*. Oxford: Oxford University Press.

Endley, M., & Karim, K. (2022). Effects of focused written feedback and revision in the development of explicit and implicit knowledge in EFL writing. *Language Teaching Research Quarterly*, *30*, 32-49. https://doi.org/10.32038/ltrq.2022.30.03

Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*, 1149-1160. https://doi.org/10.3758/BRM.41.4.1149

Fazio, L. (2001). The effects of corrections and commentaries on the journal writing accuracy of minority- and majority-language students. *Journal of Second Language Writing*, *10*(4), 235-249. https://doi.org/10.1016/S1060-3743(01)00042-X

Ferreira, A. (2017). El efecto del feedback correctivo para mejorar la destreza escrita en ELE. *Colombian Applied Linguistics Journal*, *19*(1), 37-50. http://dx.doi.org/10.14483/calj.v19n1.10220

Field, A., Miles, J., Field, Z. (2012). *Discovering statistics using R*. London: SAGE.

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

Ghazi, S., & Zamanian, M. (2016). The influence of asynchronous computer-mediated versus conventional corrective feedback on Iranian EFL learners' writing accuracy. *Journal of Studies in Learning and Teaching English*, *5*(2), 169-181.

Gray, R. (2021). Why do all systematic reviews have fifteen studies? *Nurse Author & Editor*, *30*(4), 27-29. https://doi.org/10.1111/nae2.8

Guo, Q., & Barrot, J. (2019). Effects of metalinguistic explanation and direct correction on EFL learners' linguistic accuracy. *Reading & Writing Quarterly*, *35*(2), 261-276. https://doi.org/10.1080/10573569.2018.1540320

Hernández-Sampieri, R., & Mendoza Torres, C. (2018). *Metodología de la investigación: Las rutas cuantitativa, cualitativa y mixta*. Mexico: McGraw-Hill Education. https://doi.org/10.17993/CcyLl.2018.15

Kang, E., & Han, Z. (2015). The efficacy of written corrective feedback in improving L2 written accuracy: A meta-analysis. *The Modern Language Journal*, *99*(1), 1-18. https://doi.org/10.1111/modl.12189

Kim, Y., Choi, B., Kang, S., Kim, B., & Yun, H. (2020). Comparing the effects of direct and indirect synchronous written corrective feedback: Learning outcomes and students' perceptions. *Foreign Language Annals*, *53*(1), 176-199. https://doi.org/10.1111/flan.12443

Kobayashi, H., & Rinnert, C. (1992). Effects of first language on second language writing: Translation versus direct composition. *Language Learning*, *42*, 183-215. https://doi.org/10.1111/j.1467-1770.1992.tb00707.x

Lenth, R. V. (2007). Statistical power calculations. *Journal of Animal Science*, *85*, 24-29. https://doi.org/10.2527/jas.2006-449

Liu, Q., & Brown, D. (2015). Methodological synthesis of research on the effectiveness of corrective feedback in L2 writing. *Journal of Second Language Writing*, *30*, 66-81. https://doi.org/10.1016/j.jslw.2015.08.011

Liu, Y. (2008). The effects of error feedback in second language writing. *Arizona Working Papers in SLA & Teaching*, *15*, 65-79.

Pallant, J. (2016). *SPSS survival manual: a step by step guide to data analysis using IBM SPSS* (6th ed.). New York: McGraw-Hill Education.

Park, E., Song, S., & Shin, Y. (2016). To what extent do learners benefit from indirect written corrective feedback? A study targeting learners of different proficiency and heritage language status. *Language Teaching Research*, *20*(6), 678-699. https://doi.org/10.1177/1362168815609617

Pica, T. (1983). Methods of morpheme quantification: the effect of the interpretation of second language data. *Studies in Second Language Acquisition*, *6*, 69-78. https://doi.org/10.1017/S0272263100000309

Riazantseva, A. (2012). Outcome measure of L2 writing as a mediator of the effects of corrective feedback on students' ability to write accurately. *System*, *40*(3), 1-10. https://doi.org/10.1016/j.system.2012.07.005

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

Stefanou, C., & Révész, A. (2015). Direct written corrective feedback, learner differences, and the acquisition of second language article use for generic and specific plural reference. *The Modern Language Journal*, *99*(2), 263-282. https://doi.org/10.1111/modl.12212

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

Supiani, S., Irawan, Y., Gestanti, R., & Cahyono, B. (2023). The effects of direct written corrective feedback and metalinguistic explanation on EFL students' acquisition of articles across proficiency level. *The Journal of Asia TEFL*, *20*(3), 711-719. https://doi.org/10.18823/asiatefl.2023.20.3.15.711

Tang, C., & Liu, Y. (2018). Effects of indirect coded corrective feedback with and without short affective teacher comments on L2 writing performance, learner uptake and motivation. *Assessing writing*, *35*, 26-40. https://doi.org/10.1016/j.asw.2017.12.002

Truscott, J. (1996). The case against grammar correction in L2 writing classes, *Language Learning*, *46*(2), 327-369. https://doi.org/10.1111/j.1467-1770.1996.tb01238.x

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

Van Beuningen, C. G., De Jong, N. H., & Kuiken, F. (2012). Evidence on the effectiveness of comprehensive error correction in second language writing. *Language Learning*, *62*, 1-41. https://doi.org/10.1111/j.1467-9922.2011.00674.x

Zhang, X. (2017). Reading-writing integrated tasks, comprehensive corrective feedback, and EFL writing development. *Language Teaching Research*, *21*(2), 217-240. https://doi.org/10.1177/1362168815623291

Appendix A

Studies included in the corpus

Table 1. Articles considered for the present study

Auhtor(s)	Year	Journal	Title
Bagheri, M., & Rassaei, E.	2021	*English Teaching & Learning*	The effect of two forms of written corrective feedback and ambiguity tolerance on EFL learners' writing accuracy
Diab, N.	2015	*Assessing Writing*	Effectiveness of written corrective feedback: Does type of error and type of correction matter?
Endley, M. & Karim, K.	2022	*Language Teaching Research Quarterly*	Effects of focused and written feedback revision in the development of explicit and implicit knowledge in EFL writing
Frear, D., & Chiu, Y.	2015	*System*	The effect of focused and unfocused indirect written corrective feedback on EFL learners' accuracy in new pieces of writing
Guo, Q. & Barrot. J.	2019	*Reading & Writing Quarterly*	Effects of metalinguistic explanation and direct correction on EFL learners' linguistic accuracy
Kim, Y., Choi, B., Kang, S., Kim, B., & Yun, H.	2020	*Foreign Language Annals*	Comparing the effects of direct and indirect synchronous written corrective feedback: Learning outcomes and students' perceptions
Park, E., Song, S., & Shin, Y.	2016	*Language Teaching Research*	To what extent do learners benefit from indirect written corrective feedback? A study targeting learners of different proficiency and heritage language status.
Shintani, N., & Ellis, R.	2015	*System*	Does language analytical ability mediate the effect of written feedback on grammatical accuracy in second language writing
Stefanou, C., & Révész, A.	2015	*The Modern Language Journal*	Direct corrective feedback, learner differences and the acquisition of Second Language Article use for generic plural reference
Supiani, S., Irawan, Y., Gestani, R., & Cahyono, B.	2023	*The Journal of Asia TEFL*	The effects of direct written corrective feedback and metalinguistic explanation on EFL students' acquisition of articles across proficiency level
Zhang, X.	2017	*Language Teaching Research*	Reading-writing integrated tasks, comprehensive corrective feedback, and EFL writing development