Bulletin of University of Teacher Education Fukuoka, Vol. 73, Part Ⅰ, 37—46 (2024)

Comprehensive Error Correction in English as a Foreign Language Writing: Action Research Utilizing a One-Group Pretest-Posttest Design

Kenneth BROWN

Department of English Language Education and Research

(Received September 29, 2023 / Accepted December 22, 2023)

Abstract

This study employed an action research approach to investigate the effectiveness of Written Corrective Feedback (WCF) when comprehensive error correction was employed. The pretest compositions of 25 participants were analyzed for accuracy, in terms of both grammatical and nongrammatical error types, as well as structural complexity, lexical diversity, and fluency. During the course, all errors in writing assignments were corrected and then returned to participants, who were encouraged to revise their assignments based on the feedback. Posttests were then analyzed similarly to the pretests and compared through paired samples *t* tests. Improved accuracy was found, with a larger effect size in the case of non-grammatical error types as compared to grammatical error types. Furthermore, statistically identical levels between the pretests and posttests of structural complexity, lexical diversity, and fluency suggest that WCF did not lead participants to pursue greater accuracy by avoiding more sophisticated writing. Independent samples *t* tests were conducted with the two groups that did or did not regularly revise their compositions; no differences in outcomes were found between the two groups. Implications for future teaching approaches involving WCF are discussed.

Key words: comprehensive written corrective feedback, error correction, direct and indirect feedback, accuracy development, written complexity, revisions

Introduction

If effective instruction is the first goal of the typical teacher, the immediate second goal would surely be the effective provision of feedback. In the case of L2 writing instruction, this feedback will often include Written Corrective Feedback (WCF), wherein errors are addressed in some manner. Ferris (2011) points out that most teachers "at least implicitly believe in the importance of error feedback and provide it consistently to their students" (p. 14). Zamel (1985) notes, "That writing teachers spend a great deal of time responding to their students' papers is a truism" (p. 79). No doubt the importance that teachers place on WCF is the cause of this dedication. Indeed, a survey of 1,053 L2 writing teachers in 69 different countries, conducted by Evans et al. (2010), found that an overwhelming 99% of the respondents said that they provided at least some error correction.

Overall Effectiveness of WCF

In light of the perceived importance of WCF and the considerable amount of time it takes to provide it, teachers and researchers have long been concerned with how effective it actually is. Whether WCF should even be practiced in the first place has been investigated and vigorously debated ever since 1996 when Truscott first contended that WCF, in contrast to feedback on content and clarity, should not be

practiced because it is allegedly ineffective and even potentially harmful.

There are studies that do indeed support the contention that WCF is ineffective (Kepner, 1991; Polio et al., 1998; Robb et al., 1986; Semke, 1984; Sheppard, 1992; Truscott & Hsu, 2008). However, most studies since 1996 seem to find evidence that counters Truscott's original claim (Bitchener, 2008; Bitchener & Knoch, 2008; Bitchener & Knoch, 2010a; Bitchener & Knoch, 2010b; Ellis et al., 2008; Frear, 2012; Guo, 2015; Rummel, 2014; Sheen, 2007; Sheen et al., 2009; Shintani & Ellis, 2013; Shintani et al., 2014), although the results of these studies are certainly open to interpretation (see Ferris, 2004; Truscott, 2007).

Types of WCF

Assuming WCF is effective, the next important question would be what type of WCF a teacher should give. WCF can be categorized as direct, indirect, or metalinguistic (Ellis, 2009). Direct WCF entails the teacher directly providing correct forms of words or inserting or removing words. In the case of indirect WCF, the teacher draws the writer's attention to the existence of an error, either by indicating the exact location of the error or by indicating the general vicinity, e.g., the line, where the error appears. Finally, when providing metalinguistic WCF, teachers provide explanations, often of a grammatical nature, about the errors. WCF can also be categorized by whether all errors are corrected or only a few predetermined error types. The former is referred to as comprehensive (or unfocused) WCF, and the latter is referred to as focused WCF.

Unfortunately for teachers looking for simple answers, the empirical evidence regarding types of WCF is mixed. For example, when it comes to the comparative advantages of direct and indirect WCF, some studies have found direct WCF to be more effective (Chandler, 2003; Van Beuningen et al., 2008; Van Beuningen et al., 2012), and others have found indirect WCF to be more effective (Frantzen, 1995; Lalande, 1982). Meanwhile, Ferris (2006) found that direct WCF was more effective in the short term, but indirect was more effective in the long run, i.e., when looking at accuracy gains in new texts. As for metalinguistic explanation, Bitchener and Knoch (2010b) and Guo (2015) found an advantage to this approach over the long run. In contrast, while Shintani and Ellis (2013) found an advantage to metalinguistic explanation in the short term, they did not find a long-term advantage, and Shintani et al. (2014) found it to be less effective than direct WCF.

Numerous studies have found focused WCF to be effective (Bitchener, 2008; Bitchener & Knoch, 2008; Bitchener & Knoch, 2009; Bitchener & Knoch, 2010b; Bitchener et al., 2005; Guo, 2015; Rummel, 2014; Sheen, 2007; Shintani & Ellis, 2013; Shintani et al., 2014). However, less research has been conducted on comprehensive WCF, and the results have been a little more mixed. Although Truscott and Hsu (2008) did not find comprehensive WCF effective in the long run, and Sheen et al. (2009) found it to be less beneficial than focused WCF, Van Beuningen et al. (2008, 2012) found comprehensive WCF to be effective. Ellis et al. (2008) found that both were just as effective as the other.

Revisions and Written Corrective Feedback

When it comes to the effect of revising writing that has received WCF, Truscott (1996, 2007) has stressed the important distinction between whether these revisions lead to increased accuracy in new texts or only in the already corrected composition. Among studies that have investigated this issue, Chandler (2003) found that revision following WCF not only led to greater accuracy but also fluency in the long run. Shintani et al. (2014) and Ekanayaka and Ellis (2020) found that WCF followed by revisions was even more effective than WCF without revisions when it came to improved accuracy in new writing.

Teachers' WCF Practices

As for teachers' actual WCF practices, Van Beuningen et al. (2012) suggest that comprehensive WCF is more common, although the evidence for this is inconclusive. Lee (2004, 2008, 2009) did find a pronounced preference for comprehensive WCF among 200 university and high school instructors in Hong

Kong. Similarly, in a study involving 110 primary and secondary school EFL teachers from five different countries, Furneaux et al. (2007) found a preference for direct WCF targeting grammatical errors. In contrast, in a study involving 30 EFL teachers, Salteh and Sadeghi (2015) found that a clear majority of them preferred only marking errors that interfered with effective communication, and Yang et al. (2021) found that most of the 2,012 EFL teachers in Beijing that they studied preferred focused to comprehensive WCF.

The Present Study

This study is intended as action research, which Wallace (1998) explains "is done by systematically collecting data on your everyday practice and analyzing it in order to come to some decisions about what your future practice should be" (p. 4). As a writing instructor, I have fallen into the practice of providing comprehensive WCF and strongly encouraging, but not requiring, the students to revise their corrected compositions. While these practices provide me with the sense of meeting my responsibilities as a teacher, while also at least encouraging the students to take some personal responsibility for their own improvement, the myriad issues regarding WCF make me wonder whether this approach is truly effective and worth the time and effort. I have, therefore, employed an action research approach in this study to investigate my own concerns.

This study has been somewhat inspired by that of Van Beuningen et al. (2012), in that not only am I interested in the effectiveness of comprehensive WCF but, as with Van Beuningen et al., I am concerned with some of Truscott's claims for the negative effects of WCF. Specifically, Truscott (2001, 2007) has stated that while WCF might be beneficial when it comes to non-grammatical error types, this is not the case when it comes to grammatical error types. However, Van Beuningen et al. (2012) found this to not be the case, although they did find indirect WCF to be more effective for treating non-grammatical error types and direct WCF to be more effective for grammatical error types. Truscott (1996, 2004, 2007) also says that WCF will end up discouraging students from producing more complex writing in their effort to avoid making mistakes. However, again, Van Beuningen et al. (2012) did not find this to be the case, and Fazilatfar et al. (2014) even found that comprehensive WCF had a positive effect on both syntactic and lexical complexity. At any rate, I would also like to investigate the same issues in the context of my own writing classes. I would like to point out, though, that this study cannot, by any means, be considered to be a replication of the 2012 study by Van Beuningen et al., as recommended by Bitchener and Knoch (2015), as it consists of only one group and, therefore, is lacking in experimental design.

My research questions are as follows:

RQ 1. As a teacher, can I feel reasonably confident that comprehensive WCF leads to improved accuracy in new texts?

RQ2. Does it seem that grammatical error types are less amenable to correction than nongrammatical error types?

RQ3. Does comprehensive WCF seem to lead to avoidance of more structurally complex, lexically diverse, and fluent writing?

Reflecting the cyclical nature of action research as described by Kemmis et al. (2014), I observed during the study that roughly half of the students revised their compositions more than half of the time, and roughly half of the students did not. Therefore, in an act of re-planning, I then added the following research question:

RQ 4. Is there a difference in improvement between students who choose to revise corrected compositions and those who do not?

Method

Setting and Participants

Participants were from two separate English writing classes at a university in northern Kyushu in Japan that specializes in education. The total number of participants was 25. Sixteen of those were in the class for those students seeking a secondary teaching license in teaching English. The other nine were in the class for first-year students majoring in secondary English education or special education. In the class for those seeking a secondary license, nine of the participants were in their second year, five were in their third year, and two were in their fourth year. In the other class, seven were English majors, and two were special education majors. Although the participants were from two separate classes, they were treated as one group as I applied the same method of error treatment to all of the participants.

All participants gave informed consent after being told, in Japanese, that participation was completely voluntary and that they may opt out at any time with no repercussions to themselves. Participants were also assured that the privacy of the information obtained would be properly protected and that no identifying information would be included in any reports of research findings.

Procedure

In the first class of the semester, I gave the following instructions to all students, whether participating in the study or not, for their first diagnostic writing assignment; this acted as the pretest for the participants in the study:

> If you could travel by time machine, which period of time would you visit? Give the time and your reasons.
> This is to check your English level. It is NOT for a grade. Please do not use a dictionary or the internet to help you write.
> Write about 150 words.

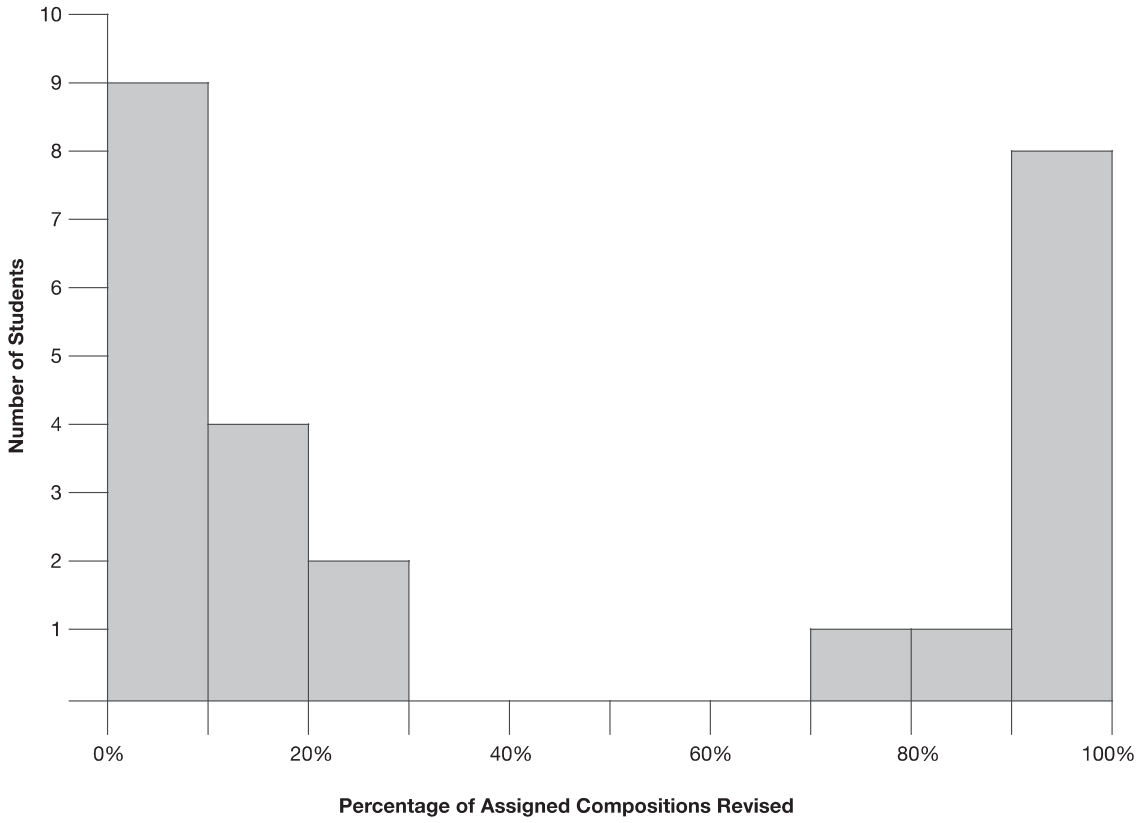
Students were allowed to write the essay on their own time at home. Despite this task not being graded, students were encouraged to do their best in order to give an accurate sample of their writing ability for diagnostic purposes. Students may have flouted the stipulation to not use a dictionary or the internet, but I hoped the explanation that this was not for a grade would eliminate the impetus to do so. I used Google Classroom to collect this task as well as the later writing assignments and the posttests.

I employed comprehensive WCF on all the subsequent assigned compositions. After returning the compositions to the students, I encouraged, but did not require, them to revise them for a higher grade. Grades were determined through the four categories of content, organization, structure, and mechanics, with each accounting for 25%. Because I had already corrected any mistakes that I found in structure and mechanics through the "edits become suggestions" function in Google Classroom, students only had to "accept" my suggested edits to correct and improve the structure and mechanics in their compositions.

Not all the students took the opportunity to revise. Of the participants, six revised 100% of their writing assignments, two revised 90% of them, one revised 86% of them, one revised 71% of them, two revised 29% of them, one revised 14% of them, three revised 10% of them, and nine did not revise any of them. (For a visual representation of this, see Figure 1.) In order to investigate the effect of revisions on long-term accuracy, I divided the students into two groups: those who revised 0–30% of their compositions, $n = 15$, and those who revised 70–100% of their compositions, $n = 10$.

Figure 1

Breakdown of Students Who Revised Their Compositions



In the last class of the semester, I gave the following instructions to all students, whether participating in the study or not, for their final writing assignment; this acted as the posttest for the participants in the study:

> If you could travel by time machine, which period of time would you visit? Give the time and your reasons.
> Write about 150 words.
> Note: You will **not** be able to rewrite this assignment.

I felt that having the topic the same as the pretest would make comparisons easier; admittedly, this does raise the possibility of a practice effect, i.e., it is possible that students might show improvement in their posttest at least partly because they already "practiced" the same task on the pretest. As with the pretest, students were allowed to write the essay on their own time at home. In this case, students were *not* discouraged from using a dictionary or the internet. This reflects my own philosophy, which I expressed in class, that all writers should not hesitate to use any tools available to them, as long as it does not result in outright plagiarism. In that sense, the intelligent use of these tools can be considered to be part of what makes a good writer. However, this must also be recognized as a possible explanation for any improvement in writing accuracy.

Analyses

I calculated overall accuracy, grammatical accuracy, non-grammatical accuracy, structural complexity, and lexical diversity of the pretests and posttests similarly to VanBeuningen et al. (2012).

Specifically, I measured the overall accuracy of both the pretests and the posttests via an error ratio: (number of linguistic errors/total number of words) × 10; because the texts were somewhat short, I used a 10-word ratio rather than the more commonly used 100-word ratio. As did VanBeuningen et al. (2012), I looked at grammatical accuracy and non-grammatical accuracy separately in order to explore Truscott's (2001, 2007) assertion that the former is less easily improved by WCF than the latter. I considered mistakes with the following to be grammatical errors: tenses, verb forms, noun forms, prepositions, articles, word order, omissions of necessary elements, additions of unnecessary elements, pronouns, and other grammatical errors. I then calculated the ratio: (number of grammatical errors/total number of words) × 10. I considered mistakes with the following to be non-grammatical errors: word choice, format, spelling, capitalization, punctuation, spacing, appropriateness, and other non-grammatical errors. I then calculated the ratio: (number of non-grammatical errors/total number of words) × 10. I measured structural complexity through a subordination index: (number of subclauses/total number of clauses) × 100. Finally, I found lexical diversity via the type-token ratio of Guiraud's Index, that is, types/ $\sqrt{}$ tokens (Guiraud, 1954, as cited in VanBeuningen et al., 2012), using Reuneker's (2017) online tool for measuring lexical diversity. To look at fluency, I straightforwardly counted the number of words written.

In preparation for comparing the two revision groups, I conducted an independent samples t test to investigate if any statistically significant differences existed between them on the pretest. On all the t tests in this study, I used a non-parametric bias-corrected and accelerated (Bca) bootstrap based on 10,000 samples, as recommended by Larson-Hall (2016). I found no significant difference between the two groups on any of the measurements of overall accuracy, grammatical accuracy, non-grammatical accuracy, structural complexity, lexical diversity, or total words written, with all the confidence intervals including zero. Throughout this study, I have chosen to interpret Cohen's d effect sizes following Plonsky and Oswald's (2014) recommendations of small (d = .40), medium (d = .70), and large (d = 1.00) in SLA research. In this case, the Cohen's d ranged from –0.52 to 0.42, all of which can be considered to be small effects.

After I evaluated the pretests and posttests, I conducted paired samples t tests to investigate any changes in accuracy and structural complexity, lexical diversity, and fluency. I used independent samples t tests to investigate any differences in the above-mentioned factors between the two revision groups.

Results and Discussion

Pretest scores, posttest scores, and the results of the paired samples t tests can be seen in Table 1.

The difference between pretest and posttest for overall accuracy of M = 0.60, SD = 0.54, BCa 95% CI [0.41, 0.80], was significant $t(24)$ = 5.53, p < .001, and represented a large effect, d = 1.11. In terms of grammatical accuracy, the difference of M = 0.28, SD = 0.39, BCa 95% CI [0.14, 0.43], was significant $t(24)$ = 3.63, p = .001, and represented a medium effect, d = 0.73. In terms of non-grammatical accuracy, the difference of M = 0.32, SD = 0.32, BCa 95% CI [0.21, 0.45], was significant $t(24)$ = 4.99, p < .001, and represented a large effect, d = 1.00.

In terms of structural complexity, the difference of M = –8.57, SD = 20.19, BCa 95% CI [–16.83, –0.89], was significant $t(24)$ = –2.12, p = .054, but represented a small effect, d = –0.43. However, the very wide confidence interval strongly suggests one should not overinterpret these results. There was no statistical difference in lexical diversity, M = –0.06, SD = 0.55, with the BCa 95% CI [–0.25, 0.14] containing zero, $t(24)$ = –0.55, p = .586, and there was practically no effect, d = –0.11. In terms of fluency (total words), the difference of M = –11.80, SD = 26.63, BCa 95% CI [–22.97, –0.83], was significant $t(24)$ = –2.22, p = .035, but represented a small effect, d = –0.44; furthermore, the confidence interval is very wide.

The results of the independent samples t tests comparing the gains on the posttest of the two revision groups can be seen in Table 2. The major takeaway, however, is that with all the confidence intervals containing zero, none of the results can be considered significant.

While it may not be possible, empirically speaking, to come to any hard conclusions, in terms of

action research, I feel I can make some reasonable suppositions. First of all, regarding RQ1, the students did show increased accuracy. While that was almost certainly not due entirely to the provided WFC, it seems unlikely to me that the WFC was entirely inconsequential. In regard to RQ2, students showed improvement in both grammatical and non-grammatical error types, although there was a larger effect size in the case of non-grammatical error types; this suggests that these error types may indeed be at least a little more amenable to correction than grammatical error types. As for RQ3, posttests did not exhibit any substantial degree of decreased structural complexity, lexical complexity, lexical diversity, or fluency in terms of total words; therefore, it seems that WCF, at least in this case, did not end up encouraging students to make their subsequent writing less sophisticated. Finally, when it comes to RQ4, it seems that whether students

Table 1

Results of Paired Samples T Tests [and BCa 95% Confidence Intervals]

	Pretest		Posttest		Difference				
Measure	M	SD	M	SD	M	SD	*t*(24)	*p*	*d*
overall accuracy	1.02 [0.83, 1.21]	0.51	0.42 [0.32, 0.54]	0.30	0.60 [0.41, 0.80]	0.54	5.53	<.001	1.11
grammatical accuracy	0.56 [0.44, 0.70]	0.35	0.28 [0.20, 0.38]	0.23	0.28 [0.14, 0.43]	0.39	3.63	.001	0.73
non–grammatical accuracy	0.45 [0.35, 0.58]	0.30	0.14 [0.10, 0.18]	0.12	0.32 [0.21, 0.45]	0.32	4.99	<.001	1.00
structural complexity	24.72 [20.65, 28.80]	10.21	33.29 [26.59, 40.54]	18.49	–8.57 [–16.83, –0.89]	20.19	–2.12	.054	–0.43
lexical diversity	6.38 [6.18, 6.59]	0.54	6.44 [6.26, 6.64]	0.55	–0.06 [–0.25, 0.14]	0.55	–0.55	.586	–0.11
fluency (total words)	151.60 [145.72, 158.12]	16.25	163.40 [157.40, 169.68]	18.05	–11.80 [–22.97, –0.83]	26.63	–2.22	.035	–0.44

Note. $N = 25$. Bootstrap results are based on 10,000 bootstrap samples.

Table 2

Results of Independent Samples T Tests for Posttest [and BCa 95% Confidence Intervals]

	Group 1: Fewer Revisions $n = 15$		Group 2: More Revisions $n = 10$		Difference				
Measure	M	SD	M	SD	M	SE	*t*(23)	*p*	*d*
overall accuracy	0.48 [0.35, 0.64]	0.32	0.34 [0.20, 0.49]	0.27	0.15 [–0.07, 0.36]	0.12	1.18	.238	0.48
grammatical accuracy	0.33 [0.23, 0.45]	0.25	0.21 [0.10, 0.33]	0.19	0.12 [–0.04, 0.28]	0.09	1.26	.202	0.52
non–grammatical accuracy	0.15 [0.08, 0.22]	0.12	0.12 [0.06, 0.19]	0.11	0.03 [–0.07, 0.12]	0.05	0.55	.578	0.22
structural complexity	37.59 [28.66, 47.77]	21.12	26.84 [18.32, 34.68]	11.87	10.75 [–0.17, 22.46]	6.50	1.46	.131	0.60
lexical diversity	6.28 [6.06, 6.51]	0.46	6.69 [6.33, 7.08]	0.60	–0.41 [–0.86, 0.03]	0.22	–1.92	.068	–0.78
fluency (total words)	167.87 [159.41, 176.63]	19.03	156.70 [148.50, 165.69]	14.92	11.17 [–2.09, 23.81]	6.64	1.56	.132	0.64

Note. $N = 25$. Bootstrap results are based on 10,000 bootstrap samples.

revised their compositions made no difference in the long run. This result was somewhat surprising to me, considering the previous research (Chandler, 2003; Ekanayaka & Ellis, 2020; Shintani et al., 2014) that suggests its effectiveness. Perhaps the fact that the groups were very much self-selected was a factor, although I still would have expected the students who revised more to show more improvement, if only because I would assume that their willingness to revise reflected higher motivation on their part. It is also possible that an effect would have been seen over a longer time span.

Conclusion

As a sort of pilot study, I feel the results here were interesting enough to warrant further research using a true experimental or quasi-experimental design that would result in generalizable data. In terms of action research and how these results will influence my future decisions as a teacher, I certainly feel that comprehensive WCF was worth taking the time and effort. Students' writing accuracy improved, and that applies to both grammatical and non-grammatical error types. At the very least, comprehensive WCF certainly did not seem to do any harm, as posttests showed no deterioration in terms of structural complexity, lexical diversity, or fluency. Nevertheless, I feel I should try other forms of WCF in addition to comprehensive WCF so as to pursue more informed decision-making on my part. I plan on continuing to encourage revisions partly because of findings in the previously mentioned studies and partly because I feel encouraging revisions is more in line with a process approach to writing. However, the fact that there was no difference between the two revision groups suggests to me that I should reconsider how I implement student revisions in my classes, while also further investigating the effect of revisions on long-term accuracy gains.

References

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

Bitchener, J., & Knoch, U. (2008). The value of written corrective feedback for migrant and international students. *Language Teaching Research*, *12*(3), 409–431. https://doi.org/10.1177/1362168808089924

Bitchener, J., & Knoch, U. (2009). The value of a focused approach to written corrective feedback. *ELT Journal*, *63*(3), 204–211. https://doi.org/10.1093/elt/ccn043

Bitchener, J., & Knoch, U. (2010a). 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., & Knoch, U. (2010b). Raising the linguistic accuracy level of advanced L2 writers with written corrective feedback. *Journal of Second Language Writing*, *19*, 207–217. https://doi.org/10.1016/j.jslw.2010.10.002

Bitchener, J., & Knoch, U. (2015). Written corrective feedback studies: Approximate replication of Bitchener & Knoch (2010a) and Van Beuningen, De Jong & Kuiken (2012). *Language Teaching*, *48*(3), 405–414. https://doi.org/10.1017/S0261444815000130

Bitchener, J., Young, S., & Cameron, D. (2005). The effect of different types of corrective feedback on ESL student writing. *Journal of Second Language Writing*, *14*(3), 191–205. https://doi.org/10.1016/j.jslw.2005.08.001

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

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., Sheen, Y., Murakami, M., & Takashima, H. (2008). The effects of focused and unfocused written corrective feedback in an English as a foreign language context. *System*, *36*(3), 353–371. https://

doi.org/10.1016/j.system.2008.02.001

Evans, N. W., Hartshorn, K. J., & Tuioti, E. A. (2010). Written corrective feedback: the practitioners' perspective. *International Journal of English Studies, 10*(2), 47–77. https://doi.org/10.6018/ijes/2010/2/119191

Fazilatfar, A. M., Fallah, N., Hamavandi, M., & Rostamian, M. (2014). The effect of unfocused written corrective feedback on syntactic and lexical complexity of L2 writing. *Procedia-Social and Behavioral Sciences, 98*, 482–488. https://doi.org/10.1016/j.sbspro.2014.03.443

Ferris, D. (2006). Does error feedback help student writers? New evidence on the short-and long-term effects of written error correction. In K. Hyland & F. Hyland (Eds.), *Feedback in second language writing: Contexts and issues* (pp. 81–104). Cambridge University Press. https://doi.org/10.1017/CBO9781139524742.007

Ferris, D. (2011). *Treatment of error in second language student writing* (2nd ed.). University of Michigan Press. https://doi.org/10.3998/mpub.2173290

Frantzen, D. (1995). The effects of grammar supplementation on written accuracy in an intermediate Spanish content course. *Modern Language Journal, 79*, 329–344. https://doi.org/10.1111/j.1540-4781.1995.tb01108.x

Frear, D. (2012). *The effect of written CF and revision on intermediate Chinese learners' acquisition of English.* [Unpublished doctoral thesis]. University of Auckland, New Zealand.

Furneaux, C., Paran, A., & Fairfax, B. (2007). Teacher stance as reflected in feedback on student writing: An empirical study of secondary school teachers in five countries. *IRAL. International Review of Applied Linguistics in Language Teaching, 45*(1), 69–94. https://doi.org/10.1515/IRAL.2007.003

Guo, Q. (2015) *The effectiveness of written CF for L2 development: A mixed-method study of written CF types, error categories and proficiency levels.* [Unpublished doctoral thesis]. AUT University, Auckland, New Zealand.

Kemmis, S., McTaggart, R., & Nixon, R. (2014). *The action research planner: Doing critical participatory action research.* Springer. https://doi.org/10.1007/978-981-4560-67-2_5

Kepner, C. G. (1991). An experiment in the relationship of types of written feedback to the development of second-language writing skills. *The Modern Language Journal, 75*(3), 305–313. https://doi.org/10.2307/328724

Lalande, J. F. (1982). Reducing composition errors: An experiment. *Modern Language Journal, 66*, 140–149. https://doi.org/10.2307/326382

Larson-Hall, J. (2016). *A guide to doing statistics in second language research using SPSS and R* (2nd ed.). Routledge. https://doi.org/10.4324/9781315775661

Lee, I. (2004). Error correction in L2 secondary writing classrooms: The case of Hong Kong. *Journal of Second Language Writing, 13*(4), 285–312. https://doi.org/10.1016/j.jslw.2004.08.001

Lee, I. (2008). Understanding teachers' written feedback practices in Hong Kong secondary classrooms. *Journal of Second Language Writing, 17*(2), 69–85. https://doi.org/10.1016/j.jslw.2007.10.001

Lee, I. (2009). Ten mismatches between teachers' beliefs and written feedback practice. *ELT Journal, 63*(1), 13–22. https://doi.org/10.1093/elt/ccn010

Plonsky, L., & Oswald, F. L. (2014). How big is "big"? Interpreting effect sizes in L2 research. *Language Learning, 64*(4), 878–912. https://doi.org/10.1111/lang.12079

Polio, C., Fleck, C., & Leder, N. (1998). "If I only had more time:" ESL learners' changes in linguistic accuracy on essay revisions. *Journal of Second Language Writing, 7*(1), 43–68. https://doi.org/10.1016/S1060-3743(98)90005-4

Reuneker, A. (2017). *Lexical Diversity Measurements.* https://www.reuneker.nl/files/ld

Robb, T., Ross, S., & Shortreed, I. (1986). Salience of feedback on error and its effect on EFL writing quality. *TESOL Quarterly, 20*(1), 83–96. https://doi.org/10.2307/3586390

Rummel, S. L. (2014). Student and teacher beliefs about written CF and the effect those beliefs have on

uptake: A multiple case study of Laos and Kuwait [Unpublished doctoral thesis]. AUT University, Auckland, New Zealand.

Salteh, M. A., & Sadeghi, K. (2015). Teachers' and students' attitudes toward error correction in L2 writing. *Journal of Asia TEFL*, *12*(3). http://dx.doi.org/10.18823/asiatefl.2015.12.3.1.1

Semke, H. D. (1984). Effects of the red pen. *Foreign Language Annals*, *17*(3), 195–202. https://doi.org/10.1111/j.1944-9720.1984.tb01727.x

Sheen, Y. (2007). The effect of focused written corrective feedback and language aptitude on ESL learners' acquisition of articles. *TESOL Quarterly*, *41*(2), 255–283. https://doi.org/10.1002/j.1545-7249.2007.tb00059.x

Sheen, Y., Wright, D., & Moldawa, A. (2009). Differential effects of focused and unfocused written correction on the accurate use of grammatical forms by adult ESL learners. *System*, *37*(4), 556–569. https://doi.org/10.1016/j.system.2009.09.002

Sheppard, K. (1992). Two feedback types: Do they make a difference?. *RELC Journal*, *23*(1), 103–110. https://doi.org/10.1177/003368829202300107

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

Shintani, N., Ellis, R., & Suzuki, W. (2014). Effects of written feedback and revision on learners' accuracy in using two English grammatical structures. *Language Learning*, *64*(1), 103–131. https://doi.org/10.1111/lang.12029

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. (2001). Selecting errors for selective error correction. *Concentric: Studies in Linguistics*, *27*(2), 93–108.

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

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

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

Van Beuningen, C. G., De Jong, N. H., & Kuiken, F. (2008). The effect of direct and indirect corrective feedback on L2 learners' written accuracy. *ITL International Journal of Applied Linguistics, 156*, 279–296. https://doi.org/10.2143/ITL.156.0.2034439

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), 1–41. https://doi.org/10.1111/j.1467-9922.2011.00674.x

Wallace, M. J. (1998). *Action research for language teachers*. Cambridge University Press.

Yang, L., Zhang, L., Li, C., Wang, K., Fan, L., & Yu, R. (2021). Investigating EFL teachers' beliefs and practices about written corrective feedback: A large-scale study. *Language Teaching Research Quarterly*, *25*, 29–65. https://doi.org/10.32038/ltrq.2021.25.03

Zamel, V. (1985). Responding to student writing. *TESOL Quarterly*, *19*(1), 79–101. https://doi.org/10.2307/3586773