**Written corrective feedback studies: approximate replication of Bitchener and Knoch (2010a) and Van Beuningen, De Jong and Kuiken (2012)**

**Abstract**

**Introduction**

The question of whether written corrective feedback (CF) has a role to play in L2 development has been controversial since Truscott (1996) published an article in *Language Learning* calling for the abandonment of the practice on theoretical, empirical and pedagogical grounds. As a result of his claims, an on-going debate about the efficacy of the practice has ensued while a number of dedicated researchers have focused their attention on empirically investigating whether learners benefit from the practice in terms of significantly improving their accuracy in subsequent texts over time. Responding not only to Truscott’s doubts about the overall effectiveness of written CF for L2 learning and his specific doubts about whether certain types of provision could ever have a meaningful and enduring effect on acquisition, further questions about the potential impact of different variables were central to the thinking and research that followed his claims. Such variables include the linguistic focus of the feedback and the relative merits of targeting (focused feedback) a limited number of error categories rather than a more comprehensive (unfocused feedback) range of error categories.While a growing body of research has begun investigating these issues over the last 15–20 years, the field is in need of replication studies before firm conclusions can be reached.

 This paper presents an argument for the replication of two studies that have investigated the key issues referred to above. The first –by Bitchener&Knoch–appeared in *Applied Linguistics* in 2010a and the second – ‘by Van Beuningen, De Jong and Kuiken–appeared in *Language Learning* in 2012. Because each of these studies investigated issues that had either not been examined beforehand or not been able to reach consistent conclusions, replication of these studies is recommended.

 Both studies have made an important contribution to the question of whether written CF can facilitate the learning, acquisition and development of second language (L2) forms and structures. Hereafter, the word ‘acquisition’ will be used to include all three terms ‘acquisition’, ‘learning’ and ‘development’. The studies are similar in that they both address controversial issues including (1) whether or not it is possible for written CF to play a role in facilitating L2 acquisition, (2) the extent to which improved accuracy (the measure of acquisition) is retained over time, (3) the effectiveness of focusing written CF on a limited or a more comprehensive range of linguistic errors, and (4) the extent to which the degree of explicitness of the types of feedback provided determines its effectiveness. In addition to also investigating these issues, the second study focused on other issues discussed in Truscott’s 1996 article, including the effectiveness of written CF for non-grammatical errors, avoidance of written complexity, and the value of written CF as a revising tool. Replications in this field are critical to the resolution of controversies about the role of written CF in L2 acquisition. Without them, conclusions will continue to be elusive because each time different variables or combinations of variables are investigated, the potential exists for such variables to mediate the findings in different ways.

 Each of these studies is able to be replicated because sufficient detail about how each was conducted has been reported. The research questions guiding both studies and the constructs and variables under investigation are clear and unambiguous. Equally, the steps involved in data collection and analysis are well explained, meaning that replication researchers would be unlikely to leave out critical steps in the collection and analysis of data. We believe that the clarity and comprehensiveness of the methodological sections of the studies will mean that findings derived from this data will be valid and reliable. This will then mean that the findings of the original studies can be appropriately verified and that appropriate generalizations can be offered.Inevitably, limitations areoften a characteristic of quasi-experimental classroom-based studies. Each of the original studies has identified what the researchers considered to be the limitations of their studies. Replication studies can therefore consider how such shortcomings might be addressed.

 The case for replicating the two original studies is presented in the following four sections. Section two provides an overview of the background of the role of written CF for L2 acquisition, explaining (1) what is controversial about the practice and (2) the current status of research evidence on the key questions that have been explored and that are recommended in this paper for replication. Sections three and four situate each study within the research context outlined in section two, provide a description of the key features of each study, their contribution to the on-going debate about the role of written CF for L2 acquisition, and methodological suggestions for their replication. Section five offers a conclusion to the case that has been presented.

**Background**

Through the decades, teachers of L2 learners and writers have believed, to some extent at least, that the practice of providing written CF on their students’ texts can help them improve the accuracy of their writing and, as a result, help them acquire L2 forms and structures that are being used incorrectly. It was not until Truscott (1996) published his critique of the practice that teachers and researchers began to ask questions about whether written CF can play a role (or indeed be expected to play a role) in helping learners acquire L2 forms and structures. His call for the abandonment of the practice led to an on-going debate on theoretical, empirical and pedagogical grounds. As a result, a steadily growing body of empirical research began to investigate key questions about (1) whether written CF can be expected to facilitate the development of improved accuracy, (2) what conditions need to be met for this to occur, and (3) whether certain pedagogical approaches are more effective than others.

 Theoretically, cases for and against the practice have been advanced in the literature. On the one hand, Krashen(1985) and Truscott (1996,1999), for example, maintain that positive evidence/input alone is sufficient for acquisition and that negative evidence/input (for example, in the form of oral and written CF) can only, at best, help develop a ‘learned’, ‘pseudo’ type of knowledge. Most frequently, this is referred to in the literature as explicit or declarative knowledge. It is generally accepted, though, that implicit knowledge (also referred to as procedural or automatized knowledge) is the ultimate goal of L2 acquisition, and debate about whether explicit knowledge can be converted to implicit knowledge through practice. Skill acquisition theories (Anderson1983, 1985; McLaughlin 1987, 1990; DeKeyser 1997, 1998) argue that explicit knowledge can be converted to implicit knowledge through practice (referred to as the interface position) whereas the non-interface position represented by Krashen and others does not. Determining whether improved written accuracy is the result of a learner drawing upon his/her explicit knowledge or implicit knowledge is a difficult task because it involves getting inside the head of the learner. However, it would seem reasonable to conclude that if improved accuracy has not resulted from implicit knowledge then, at least, the learner’s explicit knowledge has been retrieved and been the source of the improvement. In such situations, explicit knowledge can be seen to play a beneficial role in facilitating the learning process. To date, the research has tended to avoid investigations of the knowledge source of learners’ written production and focused primarily on the more pragmatically-oriented pedagogical questions about whether written CF can produce improved accuracy over time and the extent to which different factors (individual, social and contextual)might mediate a learner’s response to the feedback provided (see Bitchener&Ferris, 2012, for an overview).

 A series ofstudies (Bitchener, Young & Cameron2005;Sheen 2007; Bitchener 2008; Bitchener&Knoch 2008,2009a,b, 2010a,b; Ellis et al. 2008; Sheen, Wright &Moldawa 2009) have investigated the effectiveness of written CF for helping learners improve the accuracy of certain rule-based grammatical forms and structures (uses of the English article system; uses of the simple past tense) and the results have generally been consistent across studies. However, limited attention has been given to item-based features of the target language (e.g. the use of prepositions) and given the more idiosyncratic conditions governing their usage, more research is needed to test the effectiveness of written CF for such features. Most of the research in recent years has investigated the immediate and short term effects of providing learners with a single written CF treatment/intervention but only one study (Bitchener&Knoch, 2010a) has tested its effectiveness over a much longer period of time. Few researchers have looked at the effectiveness of providing more comprehensive written CF and of comparing targeted (focused feedback) and comprehensive (unfocused feedback) within a single research design. The research base also includes a range of studies that have investigated how explicit the written CF needs to be. Different types of written CF (indirect types such as the circling or underlining of errors; direct error correction; different types and forms of metalinguistic feedback) have been compared (see Bitchener&Ferris 2012, for an overview) but the findings have often been inconsistent, leading to a call for studies that examine the interactive effect of feedback type with other variables (e.g. proficiency level; linguistic focus of the feedback) that might refine the conditions upon which the feedback is more consistently effective. Studies comparing the relative effectiveness of a range of writing tasks and genres and of providing learners of different proficiency levels with the same written CF have yet to be reported. In addition to the focus of this research base, a growing number of studies are now investigating the mediating effect of individual, social and contextual factors on learners’ responses to the written CF they receive and the extent to which their responses play a role in learner uptake and retention (see articles in Special Issue of *Journal of Second Language Writing*2012). However, this line of research is not the focus of this paper because the two studies that we are recommending for replication did not include a focus on these factors. Rather, the two studies being recommended for replication focus on the effectiveness of different types of written CF over time in different learning contexts, on targeted and comprehensive feedback (provided on a single occasion), on the differences between two types of comprehensive feedback and the interaction effects of feedback type with linguistic structure and proficiency level. Because each of the two studies is the first to focus on a particular variable and to attempt a resolution of the inconsistent findings of earlier studiesthat had also investigated the variable, each is worthy of replication.

**Study 1 – Bitchener and Knoch (2010a)**

This study, published in *Applied Linguistics*, investigated the extent to which a single treatment/ provision of written CF was longitudinally effective for targeting two rule-based, functional uses of the English article system (the referential indefinite article ‘a’ for referring to something the first time and the referential definite article ‘the’ for referring to something already mentioned) with low-intermediate ESL learners. While earlier studies that had targeted the English article system reported consistent results, they had only investigated immediate effects (post test immediately after written CF had been given) and shorter term effects (over one or two months) of written CF. Testing Truscott’s (1996) claim that written CF should be abandoned because it is ineffective, this study investigated the claim over a ten month period. It is the only written CF study to have considered retention over such an extended period of time.

 The study also investigated the relative effect of providing learners with different types of written CF (a combination of direct error correction, written and oral meta-linguistic feedback; direct error correction and written meta-linguistic feedback; direct error correction; and no feedback for the control group). In doing so, it sought to find out whether or not the explicitness of the written CF was critical to uptake and retention. In other words, it asked whether or not greater explicitness (for example, meta-linguistic explanation and illustration as well as direct error correction) and greater quantity (for example, more than just direct error correction) would be more effective than less explicit feedback. Thus, the study sought to test the more pedagogically-oriented question (uppermost in the minds of classroom teachers) about which type of written CF has the greatest effect on improved written accuracy. The study measured effectiveness by comparing the level of improved accuracy in the immediate and delayed post-tests with that of the pre-test.Accuracy on each occasion was calculated as a percentage of correct usage on all obligatory occasions.

 Considering the longitudinal effectiveness of the written CF provided, the study found that all three treatment groups outperformed the control group in all four post-tests. This finding reveals that a single treatment of written CF on the targeted linguistic errors of low-intermediate ESL learners was effective in helping them not only improve significantly their accuracy in using the English article system for both functions immediately after receiving the feedback but that it had also been retained to the extent that they could retrieve the knowledge gained from the feedback over a ten month period. With regard to the second question, the study found no difference in effect for any of the different types of written CF provided. In other words, the most explicit feedback combination (that is, direct error correction with written and oral meta-linguistic explanation and example) was no more effective than the least explicit type of feedback (direct error correction only).

*Approach to replication*

Because this study was the first to investigate the level of accuracy retained over more than just a couple of months and because the findings show a significant level of retention as a result of providing just one written CF treatment, it is important for the field that further studies undertake a replication of the study. First, the reliability and generalizability of the study could be verified by mounting a number of approximate replications that manipulateone or more variables.

 The first approximate replication we would like to recommend is one that manipulates the population of participants. We believe thatit would be possible to access low-intermediate ESL learners from a range of backgrounds seeking to live and work in the English-speaking country where they are studying. All other variables could be kept intact. Such a study would be able to test the reliability of the original study, especially its longitudinal findings, and confirmwhether written CF has the potential to play a facilitating role in L2 acquisition. Furthermore, it would be able to test the reliability of the findings about the relative effectiveness of different types of explicit feedback. This is, arguably, the second most important issue to try to resolve given the inconsistency of earlier findings on whether greater explicitness results in greater accuracy. One of the difficulties with trying to resolve this issue is the way in which types of written CF have been operationalized in some of the earlier studies (see Bitchener and Ferris, 2012, for an overview). Some compared separate variables while others compared variable combinations, hoping to achieve greater ecological validity. Additionally, there were variations in how the variables were defined. As a result, variations such as these that are not replicated stand little chance of being tested for the reliability and generalizability of their effectiveness.

 The second approximate replication we would like to recommend is one in which the linguistic focus of the study be changed from the targeting of the English article system to the targeting of a different rule-based linguistic form or structure. further replication studies should target item-based linguistic features. It is important that other linguistic areas be targeted because the focus of most recent studies has been on the acquisition of articles. To be able to generalize the potential of written CF for L2 acquisition, other forms and structures need to be investigated.

 The extent to which written CF and its different forms can facilitate learning/improved accuracy for learners at different proficiency levels is not yet known. An approximate replication could build a comparative element into the original study, for example, by comparing learners at elementary and advanced levels. To date, studies have examined the effectiveness of written CF for targeting the English article system at both low and high proficiency levels and found it to be successful, but we do not know whether the same degree of effectiveness might be evident if other linguistic error categories are targeted. Because advanced learners have a more extensive repertoire of linguistic knowledge to draw upon, it may be that they respond to written CF more positively than lower proficiency learners when certain error categories are tested.

 A fourth recommendation for approximate replication that focuses on different learner populations and contexts would help the field understand (1) whether learners from different backgrounds (including different L1 and teaching/learning approaches) respond similarly or differently to written CF and how it is delivered and (2) whether these factors mediate uptake and retention over time. Much of the published research has focused on either ESL or EFL learners. One exception is the focus of the second study recommended for replication. In that study, the learning of Dutch was investigated. The acquisition of other target languages in both second and foreign language environments would make a worthwhile contribution to the generalizability of the two key findings of the original study. Replication researchers might consider, for example, the acquisition of Asian, European and South American languages.

Finally, we suggest two ways in which the aims of the study could be replicated conceptually. As well as the post-test writing tasks, an analysis of other writing tasks completed by the participants during class time and out-of-class time as well as in genres different to those employed in the original study would reveal the extent to which accuracy improvements in the targeted linguistic categories are found across a range of tasks types and situations, thereby providing a more holistic view of learning that has resulted from written CF.

**Study 2 – Van Beuningen, De Jong and Kuiken (2012)**

This study, which was published in *Language Learning*, set out to test several of Truscott’s (1996, 1999, 2007) controversial claims about the effectiveness of corrective feedback. Unlike the targeted, focused approach of study one, this study investigated the effect of direct and indirect comprehensive (or unfocussed) written CF on both text revisions and new pieces of writing. Investigating the effectiveness of comprehensive written feedback is important as it is a common type of feedback provided by many classroom teachers. The findings from the few studies that have tried to test this approach have been conflicting (Truscott & Hsu.2008; Van Beuningen et al. 2008;Hartshorn et al. 2010). The study also set out to investigate the interaction of these two feedback types in relation to specific error types and two different educational levels. The authors also investigated whether students avoid more complex structures because of error correction, the effect of revision without any CF, and whether time spent on CF is better spent on writing practice. Many of these aims were developed to directly address specific claims made by Truscott.The final aim of the study was to explore whether there was a possible differential effect of the two feedback types on participants from two different educational levels. The reason behind this was the hypothesis that learners from higher educational levels would have more metalinguistic awareness to incorporate indirect corrective feedback into their revisions and future new pieces of writing.

 The study employed a pre-test/single intervention/post-test (after 1 week)/delayed post-test design (after 4 weeks). The 268 participants in the study were Dutch high school students (mean age 14) learning Dutch as a second language. The writing tasks were related to biology and were collected in biology classes which aimed at integrating content and second language instruction. Apart from the two experimental groups, the study employed two control groups: one group undertook self-editing without corrective feedback and the other completed writing practice without revisions.

 The findings of the study showed that both direct and indirect comprehensive feedback resulted in improved linguistic accuracy during text revisions and in new texts (when compared to the two control groups) and that these results were durable (as measured in the delayed post-test). Both direct and indirect corrective feedback were effective, but an interaction effect with error type was found – direct feedback was more effective for grammatical errors while indirect feedback was more effective for non-grammatical errors. Finally, there was no interaction effect between the effectiveness of the corrective feedback type and the proficiency level of the learners, indicating that both groups of learners equally benefitted from both feedback types.

*Approach to replication*

The findings from Van Beuningen et al.’s study are important because they not only reject several of Truscott’s claims, but they also have direct practical applications for classroom teachers in language classrooms across the world. Van Beuningen et al.’s study was very well designed but several aspects of the study lend themselves to replication to strengthen the claims and lead to wider generalizability.

There are a number of possible directions a replication study could take. As an exact replication of this study would be difficult to accomplish, in particular considering the specific educational context the original data was collected in, we have several suggestions for approximate replication. Changing one aspect of a study and keeping all other aspects consistent is important to establish generalizability to different learning contexts. Firstly, we recommend a replication of the study with adult participants. This is important because Van Beuningen et al.’s participants were from a very specific context (high school learners of Dutch), and it is important to show that this study can be replicated with a different cohort (e.g. high school children in a different context). Similarly, it would be important to see whether this study can also be replicated with adult learners (even though this would most likely introduce two changes to the original study (i.e. age and context).If the results can be replicated with adult learners, this can have important practical implications for classroom teachers teaching adults in many contexts and language environments across the world. The findings of the initial study can then also be generalised to other contexts.

 A further suggestion for approximate replication is to investigate whether the findings are equally applicable in instructed second language as well as foreign language contexts. The participants in Van Beuningen et al.’s study were second language learners with slightly different characteristics to the two groups suggested for approximate replication above.While instructed second language learners of a language receive feedback on their language frequently and through several channels, and therefore arguably might have more input to provide them with the metalinguistic knowledge to uptake for example the indirect feedback, it is plausible that learners in a foreign language environment might be more susceptible to direct feedback. Most written CF studies to date have focussed on second language environments. It is therefore important to collect more data in foreign language classrooms to ensure generalizability of the findings.

 Our final suggestion for approximate replication is to replicate the study using a longitudinal design. Providing comprehensive written CF repeatedly over several weeks/months mirrors what learners commonly experience in classroom settings and therefore gaining a better understanding of the effectiveness of such repeated feedback is important not only theoretically but also pedagogically. To date, longitudinal feedback has mostly been employed in studies investigating the effects of focussed feedback (see Bitchener and Ferris, 2012, for overview) and it is not clear how the results from such studies transfer to contexts in which comprehensive feedback is provided. As comprehensive feedback is most commonly provided in classroom settings over several pieces of writing, this is an important gap in the literature.

 The final suggestion is for a conceptual replication of the study. Conceptual replication is a replication that addresses the same research objective as original studies but uses different methodologies (Porte 2012: 8). As a possible option for conceptual replication, we recommend adding an extra source of data to Van Beuningen et al.’s study. To gain a better understanding how students at different educational levels engage with the two types of feedback (direct vs. indirect), we recommend collecting think-aloud protocols from participants in the two experimental groups while revising texts and while composing new texts. This rich, qualitative data could provide in-depth information of how learners process written CF and might give an insight into how learning takes place. Findings from such a study could also have direct implications for theories of second language acquisition.

**Conclusion**

In conclusion, the field of written CF is an area of research in which replication studies would be important and valuable to consolidate research and make generalizations possible. This paper introduced two fairly recent original studies in this area. The studies were chosen because they were the first to investigate key questions about the efficacy of written CF and because their detailed descriptions of the methodology make replication possible. Descriptions of the design and the main findings of each of the papers were provided and a number of key proposals for replicationwere made. Findings from these studies will not only consolidate current understandings of written CF but will also be invaluable to practitioners in language classrooms. (4464 words)

**References**

Anderson, J. (1983). *The architecture of cognition.* Cambridge, MA: Harvard

University Press.

Anderson, J. (1985). *Cognitive psychology and its implications* (2nd edn). New York,

NY: Freeman.

Bitchener, J. (2008). Evidence in support of written corrective feedback.*Journal of Second Language Writing*17, 102‒118.

Bitchener, J. &D. R Ferris. (2012).*Written corrective feedback in second language acquisition and writing.*New York: Routledge.

Bitchener, J. &U. Knoch(2008).The value of written corrective feedback for migrant and international students.*Language Teaching Research*12,409‒431.

Bitchener, J., & U. Knoch. (2009a). The relative effectiveness of different types of direct written corrective feedback. *System, 37,* 322-329.

Bitchener, J., & U. Knoch. (2009b). The value of a focused approach to written corrective feedback. *ELT Journal, 63,* 204-211.

Bitchener, J. &U. Knoch(2010a). The contribution of written corrective

feedback to language development: A ten month investigation. *Applied*

*Linguistics*31,193‒214.

Bitchener, J. &U. Knoch(2010b). Raising the linguistic accuracy level of advanced

L2 writers with written corrective feedback.*Journal of Second Language*

*Writing*19,207‒217.

Bitchener, J., S. Young&D. Cameron(2005). The effective of different types of

corrective feedback on ESL student writing. *Journal of Second Language*

*Writing*14,191‒205.

DeKeyser, R. M. (1997). Beyond explicit rule learning: Automatizing second

languagemorphosyntax. *Studies in Second Language Acquisition*19, 195‒222.

DeKeyser, R. M. (1998). Beyond focus on form: Cognitive perspectives on learning and practicing second language grammar. In C. J. Doughty & J. Williams (eds.), *Focus on form in classroom second language acquisition*.New York: Cambridge University Press,42‒63.

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

Hartshorn, K. J., N. W., EvansP. F., MerrillR. R., SudweeksD. Strong-Krause, &N. J. Anderson(2010). Effects of dynamic corrective feedback on ESL writing accuracy.*TESOL Quarterly*44, 84–109.

Krashen, S.D. (1985). *The input hypothesis: Issues and implications.* London:

Longman.

McLaughlin, B. (1987). *Theories of second language learning.* London: Edward Arnold.

McLaughlin, B. (1990). Restructuring.*Applied Linguistics*11, 113‒128.

Porte, G. (2012).Introduction.In G. Porte (ed).,*Replication research in Applied Linguistics*. Cambridge: Cambridge University Press.

Sheen, Y. (2007). The effect of focused written corrective feedback and language aptitude on ESL learners’ acquisition of articles.*TESOL Quarterly*41, 255‒-283.

Sheen, Y., D. Wright&A. Moldawa(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.

Truscott, J. (1996). The case against grammar correction in L2 writing classes.

*Language Learning*46,327‒369.

Truscott, J. (1999). The case for “the case for grammar correction in L2 writing

 classes”: A response to Ferris. *Journal of Second Language Writing*8,111‒

 122.

Truscott, J. (2007). The effect of error correction on learners’ ability to write accurately.*Journal of Second Language Writing*16,255‒272.

Truscott, J., &A. Y-P.Hsu(2008).Error correction, revision, and learning.*Journalof Second Language Writing*17,292‒305.

Van Beuningen, C., N. H., de Jong&F. Kuiken(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., N. H., de Jong&F. Kuiken(2012). Evidence on the effectiveness of comprehensive error correction in Dutch multilingual classrooms.*Language Learning*62,1‒41.