Feedback Research in Second Language **EUROKD**

https://doi.org/10.32038/frsl.2023.01.03

What about validity? Thoughts on the state of research on written corrective feedback

John Truscott
National Tsing Hua University, Taiwan

Correspondence
Email: truscott@mx.nthu.edu.tw

Abstract

Validity is the most important factor in good research, and it cannot be taken for granted. As the term is used here, validity is determined by (a) whether the research actually addresses the question that it is intended to address; and (b) whether the inferences drawn from it, by its authors and by subsequent reviewers, are legitimate. Validity thus depends on the goals of the research. While a variety of meaningful goals can be identified, the primary goal of WCF research is to determine if WCF is an effective way to improve learners' ability to write accurately and should therefore be used in language instruction. Judging validity in terms of this goal, it is argued here that problems are pervasive. Goals are confused, a variety of fundamental issues in design and analysis make a study unable to address the primary question, and far too much emphasis is placed on one very narrow feature, commonly studied in ways which themselves lack validity. These problems are reflected in the meta-analyses done on this topic; as a result, this work is not presenting us with a valid picture of the effects of WCF. Altogether, this research area has major validity problems; dealing with these problems should be the top priority for researchers.

ARTICLE HISTORY

Received: 12 May 2023
Revised: 02 September 2023
Accepted: 24 October 2023

KEYWORDS

written corrective feedback, research, validity

Introduction

The efficacy of written corrective feedback (WCF) has become a major research topic in recent years, as evidenced by the establishment of a new journal dedicated (in part) to this research. This rising interest has been accompanied by an increased concern with questions of methodology and analysis and by more sophisticated uses of meta-analysis to synthesize the research findings and draw implications from them. In some respects, researchers and reviewers can no doubt take pride in the increasing sophistication of their work. But at the same time, I suggest, the most important feature of good research seems to have been lost, if not actively thrown away. This is *validity*, in its most basic sense: Does the research actually address the question(s) that it is intended to answer? Are the inferences that are drawn from it legitimate?

This is, first of all, an issue for the way that experiments are designed and analyzed. Specifically, it is about the way the design and analysis serve their goals, or fail to serve them. If they are not suitable for the question they are intended to answer, we have a validity problem. No less important is the way that research results are treated in literature reviews, especially in meta-analyses, which in principle provide the best possible picture of what has been found and what it means. If the meta-analyses are including invalid results or excluding valid results, we have a validity problem.

In WCF research, the primary goal for nearly everyone is to determine whether WCF helps learners develop their ability to write accurately and should therefore play a significant role in second language writing instruction. So the primary validity issue is whether the research and the meta-analytic syntheses of it are in fact addressing this central question. All too often, I will suggest, they are not.

Validity and the Goals of WCF Research

In judging the validity of a particular study, the first issue is what question(s) that study seeks to address. In WCF research, the primary goal is, again, to determine whether WCF helps learners develop their ability to write accurately and should therefore be used in second language writing instruction. This is the issue that I am concerned with here. But before examining the way that this issue is and is not studied, we should first take note of various other goals that might be served by WCF research.

More limited goals might be to determine if WCF is effective when done in specific ways, for specific purposes, with specific types of errors made by specific types of students (see Truscott, 1999, 2001). Kassim and Ng (2014) might be seen in these terms. When addressed to the general question of WCF's effectiveness, its special context and its narrow focus make its validity doubtful, but it does well in terms of a more modest goal of this sort (see Truscott, 2022).

Researchers might choose to assume that WCF is going to be used in some form and then ask which form is preferable. In this case a valid design simply compares the two (or more) different types of WCF. The standard requirement of a no-WCF group, essential for research on the primary question, does not apply. The findings of such research are valid as long as the authors (and, especially, later reviewers) don't try to use them to answer a different question, like whether WCF is effective.

We might also ask if WCF helps teacher trainees acquire the explicit linguistic knowledge that might be demanded of them in their future work.[1] In this case a valid test would have them explain the points that they were corrected on, or do the kinds of exercises their students will be required to do. With this type of goal there is no need for a measure of authentic writing (a basic requirement for work addressing the primary question).

[1] Whether it *should* be demanded of them is of course a very different question.

Another possible goal is to determine whether a teacher using WCF can bring about improvements in a given piece of writing without actually doing the rewriting for the student. In that case the design does not have to include anything more than a single piece of writing and the revision of that writing. We don't need to see performance on new pieces of writing – another basic requirement for addressing the primary question.

Or we can take a more theoretical approach, as seen in the currently popular practice of trying to distinguish implicit from explicit knowledge. This distinction is interesting and important for theory, and a well-developed theory can eventually serve as a guide for practice, at least in principle. But for now at least the implicit-explicit distinction does not connect to the primary, pedagogical issue. Both types of knowledge play a role in writing,[2] which is to say that valid tests of learners' ability to write accurately will encompass both types, in combination – and this is the information we need in order to address the primary pedagogical question. Exactly how each type is contributing is also an interesting question, worthy of research, but at this point at least it does not have any clear implications for the question of whether WCF improves learners' ability to write accurately or whether it should play a role in teaching. It is a theoretical goal.

Another interesting goal is to explore the in-principle effects of WCF, abstracting away from actual teaching contexts. In this case, we try to isolate its effects from those of other variables that could influence learning. Ideally, this means doing laboratory research, where rigorous controls can be imposed. When such research is carried out in actual classroom settings, it is far from clear just what things WCF should be isolated from or how this is to be done. In any case, this question is important for theory, and in the longrun, research of this sort also has potential value for teaching practice, as it can give a more fundamental understanding of the processes involved in teaching and learning. But the gap between in-principle findings and classroom application must be kept in mind. A teaching technique that looks good in isolation might well fail when placed in a more realistic context, and one that fails when used in isolation might work when supported by normal classroom activities and contexts. If the goal is to determine the value of WCF for teaching, we have to ask how well it works in the context of normal teaching practices, not how well it works in isolation from them. The pedagogical question and the in-principle question are both interesting, but they must not be confused.

While all of the above are legitimate goals, the concern in the remainder of this paper will be with the primary question, the pedagogical issue of whether WCF is useful as a general teaching tool – whether it helps learners develop their ability to write accurately. When the findings of a study are used to draw conclusions related to this primary, pedagogical question, we have to ask if the design and analysis of the study

[2] Though implicit is far more important, at least beyond the beginner level.

actually do, or can, address that question. Much too often, the answer turns out to be no.

Valid Comparisons, Part 1

It is generally accepted now that a study cannot simply look at learners' performance before and after receiving WCF and then attribute changes to the treatment, as a great many other factors could be responsible for the changes. A comparison (control) group is needed. The issue is then the proper nature of the comparison condition.

The validity of a comparison depends, again, on the goals of the research – what question(s) it is trying to answer. If the goal is to provide guidance for teaching – to tell teachers what effects WCF is likely to have in their classes – then the proper comparison is between (a) what they do in their classes if they choose to use WCF, and (b) what they do in their classes if they choose not to use it.

For (a), considerable variety is possible, and is found in practice, most notably in the type of WCF that is used. My main concern, though, is with (b), and with how it relates to (a). One possibility for a no-correction teacher is to simply return students' work to them with no comments of any kind. This means going beyond "no WCF" to "no feedback". But this approach runs counter to essentially everything that is known or believed about how writing should be taught, and has little to do with real teaching. Providing comments on what the student says – the content – is a normal practice and appears to be almost universally accepted. It is what teachers who reject WCF are likely to do and probably should do. Thus, providing the control group with feedback of this type produces a valid comparison, perhaps the most clearly valid comparison that we can have.

In much of the early research on WCF (Kepner, 1991; Semke, 1980, 1984; Sheppard, 1992) this point was recognized, or more likely taken as a background assumption, not requiring any explicit discussion or defense. These studies simply took normal writing instruction as their context, comparing the accuracy of students who received WCF to that of students who underwent the same standard instruction but without WCF. In other words, the research was well-suited to provide evidence on the value of WCF in writing instruction.[3]

But this work is now frequently dismissed (see Biber et al., 2011; Bitchener & Ferris, 2012; Kang & Han, 2015; Lim & Renandya, 2020; Sheen, 2007), or simply ignored (e.g., Brown et al., 2023; Russell & Spada, 2006). Ironically, it is dismissed *because* of its validity – because the comparison groups received content feedback rather than no feedback at all, or because they included the normal, appropriate practice of student-

[3] Additional virtues of this research were that it studied WCF as it is actually done in classes, included several treatment sessions, as is done in classes, and measured the effects by looking at authentic writing tasks. In its most important aspects, this work should be seen as a model for current researchers.

teacher conferences. The logic of this criticism (when a logic is offered) is that the effect of WCF must be isolated from other variables that might affect the success of the learning.

This criticism illustrates the problem of confusion between different goals. If the goal is to offer advice to teachers, to tell them what effects the use or non-use of WCF will have in their classes, it makes no sense to study WCF in isolation from the things that are normally, properly done in those classes. The criticism of these studies as improperly controlled is a statement to current researchers that they should avoid valid comparisons.

Judgments of validity are based, again, on the goals of the research and the conclusions that are to be drawn from it. Authors who reject the use of content feedback and conferences in the name of isolating the WCF variable are addressing a different question: the in-principle effects of WCF. This is a worthwhile question to ask, but it must be recognized that isolating the effects of WCF from what goes on in classrooms means moving away from practical issues of teaching.

This movement from the practical to the in-principle is illustrated in the comments by Bitchener and Ferris (2012, p. 56) on Sheppard's (1992) content group:

> *It is unclear whether the learners in both groups received incidental instruction on linguistic errors during the 10-week period and whether the conferences between the teacher and the content-focused group avoided any discussion of difficulties in understanding meaning that may have arisen from linguistic error. Thus, it is unclear whether in fact the content-focused group was strictly a control group.*

This might be dismissed as quibbling, but such a response would miss the real problem – the confusion between two different goals. These authors are, in principle, pedagogically oriented; using the research to draw implications for teaching appears to be their number one interest. But these comments represent a turn to the in-principle question. The claim is that in order for a no-WCF class to be a fully acceptable comparison group, we have to ensure that nothing the teacher does has any connection to the linguistic errors that students make, or else the class is not "strictly a control group". A requirement like this takes us away from the real world of teaching. A no-WCF teacher is not going to carefully self-censor to ensure that no comment on content has any relation to linguistic errors – even if this was a realistic possibility, which it is not, given the extent to which content and linguistic form are bound up with one another.

If our interest is in the pedagogical issue, the odd idea of strictly separating content comments from linguistic form is, fortunately, irrelevant. A teacher using WCF identifies linguistic errors and deliberately provides feedback on them. A teacher who

rejects WCF simply doesn't worry about linguistic errors as such, focusing instead on content. The linguistic feedback *will* have implications for content and the content feedback *will* have implications for language. So what? The pedagogical issue is not about whether a class using WCF is better than an imaginary class which strictly avoids anything that might have implications for language errors; it is about whether real WCF classes yield better results than real no-WCF classes.

Valid Comparisons, Part 2

A number of additional issues can be identified in regard to valid and invalid comparisons. The main point is that the groups should receive comparable treatment. Revision of assignments, for example, is appropriate for both types of classes, so if one group does it then the other should too. Another aspect of comparable treatment is the time that each group spends on their tasks. The less time spent on one activity – dealing with errors, for example – the more time there is for other activities. Thus, if the WCF group has been given additional tasks, such as looking over the errors that were marked, editing practice, or study of error codes, a valid comparison requires that the no-WCF group spend a similar amount of time on activities that are appropriate in a no-WCF class, probably additional reading and writing.

An issue that deserves more attention than it receives is whether the WCF has induced in learners a temporary focus on accuracy, a focus that is not shared by the comparison group and is thus a form of unequal treatment. The danger then is that corrected students might outperform uncorrected students on posttests not because they learned anything from the WCF but rather because on this task, in this context, they assign more importance to writing accurately. The point is perhaps of greatest concern when the WCF is highly focused and corrected learners have therefore received the message, implicitly or explicitly, that they should be particularly concerned with getting this point right on the posttest, a message which the comparison group has not received.

It might be argued that an increased focus on accuracy is in itself a success for WCF. This idea rests on the strong assumption that the focus will continue in the future, outside the context of the study. Many experienced language teachers can no doubt identify with Leki's (1991) observation that students who had learned to avoid certain errors in their writing class went back to making those errors immediately afterward – when filling out course evaluations. And if the focus does continue, it is far from clear that this would be a good thing, as it can be a distraction from other, possibly more important aspects of writing.

This discussion points to an interesting possibility for research – trying to induce in the comparison group the same focus that the WCF has likely induced in the experimental group. An experiment could include a WCF group, a standard no-WCF group, and an additional no-WCF group that is explicitly encouraged to focus on the targeted forms while carrying out the posttest task. A variety of results are possible,

probably depending especially on the nature of the target and the extent of the students' prior knowledge and ability with that target. In any case, the findings of such studies could have considerable value for the interpretation of WCF results in general.

Another issue that deserves attention is the attitude shown by teachers and researchers toward the treatment they are providing. What message are they giving, implicitly or explicitly, to learners in their comparison groups about the absence of WCF? In actual practice, a teacher who refrains from correcting presumably thinks that this is a good way to do things, and will share this positive attitude with students. In a valid experimental design, the no-WCF groups should receive the same positive message. At the very least, they should not be encouraged to believe that they are being deprived of valuable support. Is this validity requirement being met in WCF research? The question is unanswerable at this point, as experimental reports rarely if ever say anything about it. If a negative view of their treatment *is* being conveyed to no-WCF students, implicitly or explicitly, it constitutes a validity problem.[4]

Some Invalid Comparisons

After discussing what kinds of comparisons are valid, it is time to consider some cases in which WCF research has used comparisons that are *not* valid. It is disturbingly common in this area to see some of the most fundamental principles of experimental design and analysis disregarded, knowingly or not.

One kind of validity problem appears with surprising frequency in reviews of the research, including meta-analyses (see below). If you want to test whether Option A is better than Option B, your experiment has to compare A to B. You cannot compare A1 to A2 and then conclude that A is (or is not) better than B. But in the literature on WCF this invalid inference repeatedly appears – studies that simply compared two types of WCF to one another are presented as evidence on whether using WCF is better than not using it. The original example is Lalande (1982), who compared a traditional form of WCF to his own more elaborate version. Various authors, notably Ferris (1999), have presented the results of this comparison as evidence that WCF is effective. Also in this category is Chandler (2003), whose comparison group received WCF but was treated as a no-WCF group, with the justification that the students were not required to use the corrections until after the experiment was completed. More recent examples of this problem are the various experiments on *dynamic (written) corrective feedback*, though here the validity problem is not in the research itself but rather in the way it has been treated in the literature, notably in the meta-analyses (see below).

Another type of invalid comparison was made by Sheen et al. (2009). They used three groups: (a) a reading and writing task plus WCF; (b) the reading and writing task without WCF; and (c) neither treatment. It should be clear that if you want to judge

[4] We should also recognize the possibility of WCF students receiving a negative message about WCF, though this seems much less likely.

the effects of WCF (as the researchers did) the valid comparison is between (a) and (b), the two groups that differed only in the presence or absence of WCF. Comparison of (a) with (c) can only tell us if using the two treatments in combination is better than using neither. But the authors focused on (a) vs. (c), drawing the invalid conclusion that WCF was effective.[5] The related work of Sheen (2007, 2010) used only groups (a) and (c) and therefore had no valid comparison.

This is an example of unequal treatment being given to experimental and control groups, rendering the comparison invalid. It is certainly not the only such case. Another type of example is seen in the study by Bitchener et al. (2005),[6] in which hours of language instruction during the period of the study were 20 and 10 per week for the two WCF groups and 4 for the no-WCF group. The only valid conclusion that can be drawn from such an experiment is that intensive instruction with WCF is/is not better than non-intensive instruction without WCF.

Validity problems also arise when an experiment uses a control group that is not clearly comparable to the WCF group at the beginning of the experiment. An extreme example is Salami and Moini (2013), in which pretest scores of the experimental groups and the control group differed by several standard deviations. Findings from such a study cannot provide meaningful evidence. A less extreme case is Frear and Chiu (2015). In this case, the authors recognized the problem and adjusted their analysis accordingly, but caution still must be shown in uses of the data. In particular, it is not legitimate to use simple posttest comparisons to derive effect sizes.

In cases like these, significance testing shows that the groups were not initially comparable. What about cases in which the differences between groups are not statistically significant? The standard assumption is that we can disregard such contrasts and confidently proceed with the experiment, attributing posttest differences to the treatment (WCF). But despite the almost universal acceptance of this practice, it is not valid. We can hardly imagine an author saying "There is only a 90% chance that the groups are different, so we can safely assume they are not different", but this is the kind of reasoning that underlies this use of significance testing. The absence of a significant difference does not tell us that real differences do not exist, or that they probably do not exist, or anything of the sort. It says we cannot be sure that they do exist, which is an entirely different matter. A finding that the groups *might* be comparable – we're not sure that they aren't – is not a solid basis for carrying out an experiment or for interpreting its results.

There is no simple, general cure for this problem, at least none that I know of, but improvements can be made. We need to examine the pretest differences and make a rational, if necessarily uncertain, judgment of whether they are meaningful. This

[5] The point is particularly important because the valid comparison, A vs. B, yields the opposite conclusion.

[6] Note that this was not an English article study (to be discussed below).

means looking at the group means, asking by what fraction of a standard deviation they differ. It means looking at the p value – the farther it is above .05, the more confident we can be. In some cases it is possible to *make* the groups more similar (before the treatment, crucially) by switching some participants between groups or removing some from one or both, based on pretest scores. While comparability is not a simple yes or no matter, it is possible to make rational judgments and, sometimes, adjustments. We have to get beyond the comfortable fiction that the absence of statistical significance means everything is okay.

Perfect comparability between groups is an ideal, one that can be impractical in the real world of WCF research. But most of the problems described here are not difficult to avoid. Those that cannot be avoided can be minimized. When meaningful problems remain despite a researcher's best efforts, the resulting limitations must be recognized, both by the original author and by those who later offer reviews of the work.

The Appropriateness of the Measures Used

Some authors studying the effects of WCF have relied on tests that are essentially grammar exercises, with no claim to validity as measures of learners' writing ability.[7] There appears to be a general recognition now of this validity problem, and so it has declined considerably in recent years (for discussion of the older cases, see Truscott, 1996). But it has not disappeared. Hosseini (2015) used fill-in-the-blank in isolated sentences. The posttest used by Ahmadi et al. (2012) was (a) change adjective clauses to participle phrases, and (b) find and correct particular types of errors in a text. Even on this well-known and widely accepted point, there is still a need for greater concern with validity.

Some researchers have used measures of learners' writing ability which, while representing improvements on grammar exercises, are at best questionable in terms of validity. These measures, dictogloss and text reconstruction, have learners listen to or read some material and then write it down from memory (e.g., Frear & Chiu, 2015; Shintani et al., 2014). The use of additional measures like these no doubt adds sophistication to the research, but if you want to know how accurately students can write, it is greatly preferable (to put it mildly) to look at their actual writing. Every step away from authentic writing is a step away from validity. Interestingly, the authors of two recent meta-analyses (Brown et al., 2023; Reynolds & Kao, 2022) have incorporated this point in their criteria for including or excluding studies (but see discussion of these meta-analyses below).

In some cases at least, the use of these doubtful measures was based on feasibility. The particular aspect of form that was to be targeted by the WCF would not occur

[7] They may well have validity as measures of learners' explicit grammar knowledge. The (familiar) point is simply that they must not be treated as evidence on a question which they cannot address – whether WCF helps learners develop the ability to write accurately.

frequently enough in students' natural writing and so the validity-challenged measures were selected in order to force their use. This is sometimes the reality of experimentation – sacrifices have to be made in order to carry out the research. But while at times we may have to accept the sacrifices in validity, we do not have to pretend that there is no problem. The validity issue has to be recognized in any discussion or use of the information that is produced by such measures.

Finally, measures that might in themselves be valid can be rendered invalid by an inappropriate choice of analysis. The *suppliance in obligatory contexts* (SOC) analysis, used for example by Rummel and Bitchener (2015) in their study of English past tense, is the most important example of this problem. The limits of this measure have been recognized for a very long time (see Larsen-Freeman & Long, 1991). Using a form accurately means both using it in places where it is required and *not* using it in places where it does not belong. But SOC looks only at the former. Thus, it is more a measure of *frequency* of use than of *accuracy* of use.

WCF as a Teaching Technique vs. WCF as a Means of Improving a Written Product

One validity problem that has been with us for some time is the confusion between the main question – whether WCF helps learners develop their ability to write accurately – and the entirely distinct question of whether a particular product can be improved by learners if they use teachers' markings on that product. The latter has been the topic of a few experiments (Ashwell, 2000; Fathman & Whalley, 1990; Ferris & Roberts, 2001) and one part of several additional studies (e.g., Karim & Nassaji, 2020; Khanlarzadeh & Nemati, 2016; Sachs & Polio, 2007; Shintani & Ellis, 2013; Truscott & Hsu, 2008). The answer is yes, students do have some ability to use written corrections on their assignment to revise that assignment. But this unsurprising result does not tell us anything about whether WCF improves learners' ability to write accurately. It is addressing an entirely different question and should never be treated on a par with studies that do address the main question. It should absolutely not be averaged in with the results of those studies, as combining the two types yields numbers that are not a valid measure of anything. Most researchers seem to recognize this point now, but the confusion is still with us, as will be seen below.

Mixing First Language and Second Language Instruction

It is generally recognized that teaching second language learners is not the same thing as teaching native speakers. Not surprisingly, then, almost all discussion and debate on the value of WCF has assumed a strict separation between first and second language instruction. In discussions of the latter, first language studies are occasionally noted in passing but are given little or no weight. The striking (and puzzling) exception is the work of van Beuningen et al. (2008, 2012), which is routinely treated as evidence on the value of WCF in second language instruction, which it is not.

The target language in this work, carried out in the Netherlands, was Dutch. Most of the participating students were born in the Netherlands and all had used Dutch in

school at least since age 4 (i.e., for ten years or more). Of these students, 24 came from families in which Dutch was the *only* language used at home. Setting these 24 aside, the authors in effect said that they had no idea how many of their participants were exposed to Dutch from birth or to what extent they had been exposed. Nor were they concerned about whether participants were classified as native or non-native.[8] The students were considered suitable for the studies simply because they all had "a relatively low level of writing proficiency" (van Beuningen et al., 2012, Note 1). These were not studies of second language learning, and their findings cannot be validly presented as evidence on the effects of WCF in second language instruction.

Strong Results at the Expense of Validity: The English Article Studies

The proper use of English articles can be quite challenging for language learners, and its importance is clear, so it constitutes a natural target for WCF research. The complexity of the article system should raise strong doubts, though, about whether general improvements through WCF are feasible. Perhaps for this reason, Bitchener (2008; Bitchener & Knoch, 2008, 2010a, 2010b) chose to focus on one specific aspect of the system and see if improvements could be brought about in that aspect. This meant looking specifically at the use of *a* for first use ("I met *a* man this morning") vs. *the* for subsequent mentions ("*The* man was a teacher"). To carry out the experiment, the authors designed writing tasks to support the focus on this aspect of grammar, maximizing the contexts in which the "first use vs. subsequent mention" rule would dictate the use of the articles. The posttests were the same type of task. The analysis was suppliance in obligatory contexts; in other words, it looked exclusively at the use of the articles in contexts in which this rule dictated their proper use.

This highly focused approach yielded strong positive results. But the strong results came at the expense of validity.

The most obvious limitation of this research is how narrow it is, with the implication that the findings have little to say about the general effectiveness of WCF or whether it should be used by teachers. It would be a mistake (one that sometimes appears in the literature) to say simply that this work studied the correctability of "articles", as the research picked out only the one, simplest aspect of this very complex system. The narrowness of the treatment also raises the question of learners' attention during the testing. Learners who were corrected specifically on one point and then asked to do the same kind of task in the same context are quite likely to be giving that point more attention than they otherwise would, possibly inflating the results.

But the most important validity problem with this work is the way it isolated the target rule from the real-life complexities of article use. The use of English articles is based on a very complex system, of which the target rule is only one part. Whether a

[8] Given that Dutch had been their school language from age 4, it is not clear that *any* of them can be properly classified as second language learners.

particular article should be used in a given case is determined by interactions among the various rules, occurring – crucially – within specific contexts. Learning to use English articles properly means developing this system, learning in what contexts a rule should be applied and when it should give way to other rules.

The Bitchener experiments did not look at any of this. Instead, they selected the simplest aspect of the system and removed it from the system as a whole, designing tasks specifically for the use of the target rule. To get a high score on these tasks, learners only needed to blindly apply that rule. The tests offered them little opportunity to show that they could judge when it should or should not be used. For this reason alone, the results cannot be considered valid evidence on the primary pedagogical question.

But it gets worse. The analysis took this narrow focus to its logical conclusion, by simply discarding all cases in which proper article use was not dictated by the target rule (see the discussion of SOC above). Learners need to know not only when to apply a rule but also when *not* to apply it, but the analysis ensured that this crucial knowledge (or its absence) would play no role in the results or conclusions of the study. A learner who blindly applied the target rule, everywhere, could get not just a high score but a perfect score, even if they had no idea when this rule should or should not be used – even if the WCF had actually *harmed* the article system by getting them to focus on this one rule at the expense of everything else.

Of course we don't know that this happened. In fact, we don't know much of anything about what actually happened in these experiments. The design and analysis make it impossible to know how the WCF affected learners' knowledge of articles or, in particular, their ability to use them in normal contexts. In other words, the studies did not address the primary pedagogical issue.

The problem goes well beyond these three or four experiments.[9] This work, along with the related research of Sheen (2007, 2010; Sheen et al., 2009), has inspired a large number of additional studies. While variability exists among them in their design and analysis, and therefore in the validity issue, validity is a concern all through this set of studies. What they all share, of course, is an extremely narrow focus, with the implication that, even if other concerns are set aside, they have little to tell us about the overall effectiveness of WCF or about whether it should serve as a general teaching tool. I will return to this point shortly.

[9] I have never been sure how many distinct studies were actually done. The various meta-analyses recognize three (Brown et al., 2023) or four (Kang & Han, 2015; Lim & Renandya, 2020); I am inclined to go with three. Confusion occurs because the results were presented in at least six different papers, and the authors were not entirely forthcoming on matters of overlap and repetition.

Meta-analyses

Validity is a concern not only for the design and analysis of research but also for what is done with the findings of the research, particularly the information and advice offered to teachers. Meta-analysis offers what is probably the best way to synthesize and summarize the findings, and a number of meta-analyses have been done on WCF. This section will offer a brief survey of the meta-analyses and their issues with validity. But first, one line of research – the English article studies, and especially Bitchener's contributions – deserves special attention due to the oversized role it has played.

The Centrality of the English Article Studies

Starting, again, with the Bitchener research, it should be clear from the above discussion that these studies have major, even overwhelming, validity problems. Given these problems, their findings have little if anything to tell us about the question of WCF's general effectiveness or whether it should be a part of writing instruction. But this has not prevented the work from acquiring a prominent place in reviews of WCF research.

All of these studies have found their way into the meta-analyses. In Lim and Renandya (2020), this work contributed the highest effect size and two others that were among the highest. Kang and Han's (2015) sample had three studies that yielded large effect sizes – all three came from Bitchener's work on English articles. This work also plays a substantial, if smaller, role in the meta-analysis of Brown et al. (2023). The prominence of these studies in the meta-analyses is, in itself, a validity problem.

But the problem goes beyond the treatment of Bitchener's research. As described above, that research has inspired a great many other studies on English articles, typically focused on the same simple rule. Because of this very narrow focus, these studies should, at least, be recognized as a distinct body of research and not simply merged with other work.

As anyone familiar with the literature knows, however, this is not the way they are treated. It is little exaggeration to say that this work has, implicitly or explicitly, dominated the discussion of WCF's effectiveness. While Bitchener's work has been repeatedly cited, for many years now, as a demonstration of the effectiveness of WCF, a similarly disturbing problem is that the sheer number of these studies has given them a central role in the meta-analyses. Brown et al. (2023) includes, by the authors' count, 16 studies, or 1/3 of the sample, that reported effects of WCF on articles. In the meta-analyses by Kang and Han (2015) and Lim and Renandya (2020), the proportion is somewhat higher. In Kao and Wible's (2011) meta-analysis, the point of which was to separate the effects of focused and unfocused WCF, the "focused" half consisted entirely of studies in this group, a situation that was improved only slightly in the extended version of the meta-analysis (Kao & Wible, 2014).

When general inferences about the efficacy of WCF – and about what should be done in the classroom – rely so heavily on a single line of work, we have a validity problem. When that work has the limitations of the English article research, we have a very big validity problem.

The Meta-analyses: A Brief Survey

While the prominence of the English article studies is probably the most serious validity problem for the meta-analyses, it is by no means the only problem. The discussion here will be brief, simply highlighting a few validity problems that appear. I have discussed some of the meta-analyses in more detail elsewhere (Truscott, 2016, 2020, 2022) and will address others in the future.

Russell and Spada (2006)

This was actually a combination of written and oral research. The written side consisted of one study that looked at learners' ability to identify and correct errors that had been implanted in a news story (Lee, 1997), two revision studies measuring improvements in a particular product after learners received WCF on that product (Fathman & Whalley, 1990; Ferris & Roberts, 2001), and two studies of original writing – Fazio (2001), which obtained a negative effect that was reported as positive in the meta-analysis, and Kepner (1991). The reliance on studies that do not provide information on learners' writing ability is in itself a validity problem. The greater problem, though, is that this meta-analysis is so often uncritically cited as evidence on WCF's effectiveness (e.g., Brown et al., 2023; Kang & Han, 2015).

Biber et al. (2011)

This was a combination of L1 and L2 work. The authors also chose to include revision studies and peer feedback research, along with several studies identified above as lacking in validity (Bitchener et al., 2005; Chandler, 2003; Lalande, 1982; Lee, 1997). They explicitly excluded studies that incorporated standard classroom practices like conferences, in the interest of isolating variables. This meta-analysis is not useful for the question we are interested in here.

Kao and Wible (2011, 2014)

As described above, these authors sought to separate the effects of focused and unfocused WCF. The important point is the reliance on English article studies. In the original version the "focused" group was simply these studies. The problem was not that relevant studies were excluded – there was very little material to work with at the time. It was, rather, the presentation of this validity-challenged research simply as "focused grammar feedback", contributing to what is probably the field's most significant validity problem. In the extended, 2014, version, the focused group was still made up overwhelmingly, though now not exclusively, of English article studies; among them were four Bitchener studies and Sheen (2007).

Kang and Han (2015)

As noted above, the English article studies played a very large role in this meta-analysis, creating a very large validity issue. Also included were dynamic corrective feedback studies that simply compared two types of WCF (Evans et al., 2011; Hartshorn et al., 2010). These researchers were explicit about the nature of their comparison groups: "traditional writing classes" in which learners were given "a wide variety of feedback on the linguistic accuracy of what they produced" (Hartshorn et al., p. 95). If the question is whether one type of WCF does better than another, such comparisons are appropriate, but when the question is whether correcting is better than not correcting (as it was for Kang and Han), this work is not relevant. The meta-analysis also included the mixed first and second language study of van Beuningen et al. (2008) as well as a few other studies that lacked a valid control condition (Bitchener et al., 2005; Chandler, 2003; Sheen, 2007). It excluded some relevant studies because they included standard classroom practices, in the name of isolating variables.

Lim and Renandya (2020)

Again, a large proportion of the studies used in this meta-analysis targeted English articles, with a large influence on the overall effect size. Another of them (Hosseini, 2015) used as its measure fill-in-the-blank exercises in isolated sentences, as described above. Results of the invalid tests used by Ahmadi et al. (2012), also described above, were also included in the meta-analysis. The sample included three dynamic corrective feedback studies that simply compared one type of WCF to another type (Evans et al., 2011; Hartshorn & Evans, 2015; Hartshorn et al., 2010), along with a revision study (Ferris & Roberts, 2001) and, apparently, the data from a revision task in another study (Khanlarzadeh & Nemati, 2016). Sheen (2007, 2010), with the invalid control conditions described above, were included, as was the work of van Beuningen et al. (2008, 2012), which included large numbers of native speakers. One of the other included studies (Shintani et al., 2014) relied on dictogloss as its measure and another (Frear & Chiu, 2015) used text reconstruction. The authors also removed early studies from the sample, partly because their no-WCF groups received feedback on content.

Reynolds and Kao (2022)

This meta-analysis specifically targeted unfocused WCF, so the English article studies played no role, but a number of other validity issues are present. The validity of conclusions about unfocused WCF is inevitably limited by the inherent vagueness of "unfocused". Reynolds and Kao defined it as correction of "all or most error types" rather than "one or a few" (p. 7), but it is often unclear which side of this divide a study should be placed on. Farrokhi and Sattarpour (2011, 2012), for example, corrected five error types, while the "unfocused" group of Ellis et al. (2008) received two corrections of the target error and "up to two" of other types; both were included in the meta-analysis. Complications also arise with cases in which a variety of errors were corrected but the testing was restricted to one type (e.g., Araghi & Sahebkheir, 2014; Jiang & Xiao, 2014). Three studies that relied on text reconstruction tasks (Frear, 2010; Frear

& Chiu, 2015[10]; Pashazadeh, 2017) were included, contrary to the stated policy of using "only studies in which participants composed essays" (p. 8). Al-Hazzani and Altalhab (2018) appears to be an extended revision study, though the authors' descriptions leave considerable doubt as to what they actually did. As described above, Salami and Moini's (2013) experiment began with an enormous difference between the no-WCF group and the WCF groups, rendering the posttest comparisons (and therefore the effect sizes) meaningless. Also inappropriately included were van Beuningen et al. (2008, 2012).

Brown et al. (2023)

The number one issue here, again, is the predominance of article studies, constituting a disturbing 1/3 of the sample. There are also various issues with valid measurement and comparisons. The authors made the commendable decision to include only studies that used "free writing", explicitly excluding dictogloss measures, for example. But they nonetheless included Frear and Chiu (2015), in which the measure was reconstructing a text after reading and studying it – the written version of dictogloss. The summarizing task of Nemati, Alavi, and Mohebbi (2019), which contributed by far the highest effect size in the sample, is another doubtful example of free writing. The meta-analysis also included dynamic corrective feedback studies (Evans et al., 2011; Hartshorn, 2008; Hartshorn & Evans, 2015), which, again, simply compared different types of WCF. Also included were the invalid comparisons of Sheen (2007, 2010) and Bitchener et al. (2005). The inclusion of Rummel and Bitchener (2015), with its reliance on an invalid analysis (see above), constitutes an additional validity problem. And, again, van Beuningen et al. (2008, 2012) were also inappropriately included. Excluded without explanation were Kepner (1991), Semke (1984), and Sheppard (1992).

I have left out my own humble meta-analysis (Truscott, 2007) because I find no validity problems there. It has often been criticized for its small sample size, which was a product of my decision to include only studies that validly addressed the primary pedagogical question (and also because the research available at the time was relatively limited).

Conclusion

As part of their meta-analysis, Brown et al. (2023)[11] examined a variety of factors that might be expected to influence the success of WCF. This examination showed some interesting trends in the data: requiring students to revise their work based on the feedback they receive is harmful; having learners review the corrections they receive is harmful; training in how to respond to feedback is harmful; when a particular form is targeted for WCF, instruction on that form is unhelpful, and possibly harmful;

[10] Note also that pretest contrasts invalidate any direct use of posttest contrasts in effect size calculations, which seems to be what Reynolds and Kao did.

[11] See also the supplemental material for this article: https://izeh.github.io/meta.html#6_Publication_bias

general instruction on grammar is harmful. Most strikingly: Correcting once is more effective than correcting several times; when feedback is provided more than once, the additional feedback partially reverses the benefits of the initial corrections. To be clear, this is not a matter of "diminishing returns" (see the Discussion section of Brown et al.) but rather of *negative* returns – the additional feedback is harmful.

Probably no one thinks this list accurately reflects reality. Nor is anyone likely to see the items as useful guidelines for teaching. But they *are* accurate descriptions of the findings from these 50 studies on WCF. What this suggests is that something is wrong in this body of research – it is not presenting us with a valid picture of the effects of WCF.

This is the message I have sought to convey in this paper – that things are not as they should be in the world of WCF research. While the reasons for these specific trends are far from clear, it *is* clear that validity issues pervade the work on this topic and so we should not be surprised when the findings point to implausible conclusions. There is a serious validity problem in WCF research, and this problem should be the number one concern both for researchers and for those seeking to apply their findings.

In addressing the problem, the first step is to get serious about one simple question: Why are we doing this? Various answers are possible, and they must be clearly, explicitly distinguished from one another. Once the goals of a project have been clearly established, we need to ensure that everything that is done in the project serves those goals. While this advice may sound trivial – choose a goal and stick to it – the problems identified here show that it is not. Meeting the requirements for valid research is in fact a challenging task. But if we are to obtain valid answers to the questions we are interested in, it is an essential task.

ORCID

 https://orcid.org/0000-0003-2107-567X

Acknowledgements
Not applicable.
Funding
Not applicable.
Ethics Declarations
Competing Interests
No, there are no conflicting interests.

References

Ahmadi, D., Maftoon, P., & Mehrdad, A. G. (2012). Investigating the effects of two types of feedback on EFL students' writing. *Procedia – Social and Behavioral Sciences*, *46*, 2590–2595.

Al-Hazzani, N., & Altalhab, S. (2018). Can explicit written corrective feedback develop grammatical and lexical accuracy of Saudi EFL learners? *International Journal of Education & Literacy Studies, 6*(4), 16-24. https://doi.org/10.7575/aiac.ijels.v.6n.4p.16

Araghi, S. M., & Sahebkheir, F. (2014). The effect of focused versus unfocused corrective feedback on developing grammatical accuracy of Iranian EFL learners' written performance within different gender groups. *Indian Journal of Fundamental and Applied Life Sciences, 4*, 720-729.

Ashwell, T. (2000). Patterns of teacher response to student writing in a multiple-draft composition classroom: Is content feedback followed by form feedback the best method? *Journal of Second Language Writing*, *9*, 227–257.

Biber, D., Nekrasova, T., & Horn, B. (2011). *The effectiveness of feedback for L1-English and L2 writing development: A meta-analysis*. TOEFL iBT Research Report No. TOEFLiBT-14. Educational Testing Service.

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

Bitchener, J., & Knoch, U. (2008). The value of written corrective feedback for migrant and international students. *Language Teaching Research, 12*, 406-431. https://doi.org/10.1177/1362168808089924

Bitchener, J., & Knoch, U. (2010a). The contribution of written corrective feedback to language development: A ten month investigation. *Applied Linguistics, 31*, 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., Young, S., & Cameron, D. (2005). The effect of different types of feedback on ESL student writing. *Journal of Second Language Writing, 14*, 191-205. https://doi.org/10.1016/j.jslw.2005.08.001

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

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., 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*, 353-371. https://doi.org/10.1016/j.system.2008.02.001

Evans, N. W., Hartshorn, K. J., & Strong–Krause, D. (2011). The efficacy of dynamic written corrective feedback for university-matriculated ESL learners. *System*, *39*, 229–239. https://doi.org/10.1016/j.system.2011.04.012

Farrokhi, F., & Sattarpour, S. (2011). The effects of focused and unfocused written corrective feedback on grammatical accuracy of Iranian EFL learners. *Theory and Practice in Language Studies, 1*, 1797-1803. https://doi.org/10.4304/tpls.1.12.1797-1803

Farrokhi, F., & Sattarpour, S. (2012). The effects of direct written corrective feedback on improvement of grammatical accuracy of high-proficient L2 learners. *World Journal of Education*, *2*(2). https://doi.org/10.5430/wje.v2n2p49

Fathman, A. K., & Whalley, E. (1990). Teacher response to student writing: Focus on form versus content. In B. Kroll (Ed.), *Second language writing: Research insights for the classroom* (pp. 178–190). Cambridge University Press.

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

Ferris, D. (1999). The case for grammar correction in L2 writing classes: A response to Truscott (1996). *Journal of Second Language Writing*, *8*, 1–11.

Ferris, D., & Roberts, B. (2001). Error feedback in L2 writing classes: How explicit does it need to be? *Journal of Second Language Writing*, *10*, 161–184.

Frear, D. (2010). The effect of focused and unfocused direct written corrective feedback on a new piece of writing. In S. C. Huang (Ed.), *Selected papers from the Third Conference on College English* (pp. 57-71). Taipei, Taiwan: Foreign Languages Center, National Chengchi University.

Frear, D., & Chiu, Y.-h. (2015). The effect 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

Hartshorn, K. J. (2008). *The effects of manageable corrective feedback on ESL writing accuracy* (Doctoral dissertation). Retrieved from https://scholarsarchive.byu.edu/etd/1522

Hartshorn, K. J., & Evans, N. W. (2015). The effects of dynamic written corrective feedback: A 30-week study. *Journal of Response to Writing, 1*(2), 6-34.

Hartshorn, K. J., Evans, N. W., Merrill, P. F., Sudweeks, R. R., Strong–Krause, D., & Anderson, N. J. (2010). Effects of dynamic corrective feedback on ESL writing accuracy. *TESOL Quarterly*, *44*, 84–109. https://doi.org/10.5054/tq.2010.213781

Hosseini, S. B. (2015). Written corrective feedback and the correct use of definite/indefinite articles. *International Journal on New Trends in Education and Their Implications*, *6*(4), 98–112.

Jiang, L., & Xiao, H. (2014). The efficacy of written corrective feedback and language analytic ability on Chinese learners' explicit and implicit knowledge of English articles. *English Language Teaching*, *7*(10), 22–34. https://doi.org/10.5539/elt.v7n10p22

Kang, E., & Han, Z. (2015). The efficacy of written corrective feedback in improving L2 written accuracy: A meta-analysis. *Modern Language Journal, 99*, 1-18. https://doi.org/10.1111/modl.l2189

Kao, C.-W., & Wible, D. (2011, October). *The distinction between focused and unfocused grammar feedback matters: A meta-analysis*. Paper presented at the Second Language Research Forum, Ames, IA.

Kao, C.-W., & Wible, D. (2014). A meta-analysis on the effectiveness of grammar correction in second language writing. *English Teaching & Learning*, *38*(3), 29-69. https://doi.org/10.6330/ETL.2014.38.3.02

Karim, K., & Nassaji, H. (2020). The revision and transfer effects of direct and indirect comprehensive corrective feedback on ESL students' writing. *Language Teaching Research, 24*, 519-539. https://doi.org/10.1177/1362168818802469

Kassim, A., & Ng, L. L. (2014). Investigating the efficacy of focused and unfocused corrective feedback on the accurate use of prepositions in written work. *English Language Teaching*, *7*(2), 119–https://doi.org/130. 10.5539/elt.v7n2p119

Kepner, C. G. (1991). An experiment in the relationship of types of written feedback to the development of second-language writing skills. *Modern Language Journal, 75,* 305-313.

Khanlarzadeh, M., & Nemati, M. (2016). The effect of written corrective feedback on grammatical accuracy of EFL students: An improvement over previous unfocused designs. *Iranian Journal of Language Teaching Research, 4*(2), 55-68.

Lalande, J. F. II. (1982). Reducing composition errors: An experiment. *Modern Language Journal, 66,* 140-149.

Larsen-Freeman, D., & Long, M. H. (1991). *An introduction to second language acquisition research*. Longman.

Lee, I. (1997). ESL learners' performance in error correction in writing: Some implications for teaching. *System, 25*, 465–477.

Leki, I. (1991). The preferences of ESL students for error correction in college-level writing classes. *Foreign Language Annals, 24,* 203–218.

Lim, S. C., & Renandya, W. A. (2020). Efficacy of written corrective feedback in writing instruction: A meta-analysis. *TESL-EJ, 24*(3), 1–26.

Nemati, M., Alavi, S. M., & Mohebbi, H. (2019). Assessing the effect of focused direct and focused indirect written corrective feedback on explicit and implicit knowledge of language learners. *Language Testing in Asia, 9*(7). https://doi.org/10.1186/s40468-019-0084-9

Pashazadeh, A. (2017). The effect of mid-focused and unfocused written corrections on the acquisition of grammatical structures. *Journal of Response to Writing, 3,* 56-82.

Reynolds, B. L., & Kao, C.-W. (2022). A research synthesis of unfocused feedback studies in the L2 writing classroom: Implications for future research. *Journal of Language and Education, 8*(4), https://doi.org/5-13. 10.17323/jle.2022.16516

Rummel, S., & Bitchener, J. (2015). The effectiveness of written corrective feedback and the impact Lao learners' beliefs have on uptake. *Australian Review of Applied Linguistics, 38,* 66–84. https://doi.org/10.1075/aral.38.1.04rum

Russell, J., & Spada, N. (2006). The effectiveness of corrective feedback for the acquisition of L2 grammar: A metaanalysis of the research. In J. M. Norris & L. Ortega (Eds.), *Synthesizing research on language learning and teaching* (pp. 133–164). Benjamins.

Sachs, R., & Polio, C. (2007). Learners' uses of two types of written feedback on a L2 writing revision task. *Studies in Second Language Acquisition, 29,* 67–100. https://doi.org/10.10170S0272263107070039

Salami, M., & Moini, M. R. (2013). The impact of indirect focused and unfocused corrective feedback on written accuracy. *International Journal of Foreign Language Teaching and Research, 2,* 32-41.

Semke, H. M. (1980). *The comparative effects of four methods of treating free-writing assignments on the second language skills and attitudes of students in college level first year German* (Doctoral dissertation). ProQuest Dissertations and Theses.

Semke, H. D. (1984). Effects of the red pen. *Foreign Language Annals, 17,* 195–202.

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. (2010). Differential effects of oral and written corrective feedback in the ESL classroom. *Studies in Second Language Acquisition, 32,* 201–234. https://doi.org/10.1017/S0272263109990507

Sheen, Y., Wright D., & Moldawa, A. (2009). Differential effects of focused and unfocused written correction on the accurate use of grammatical forms by adult ESL learners. *System*, *37*, 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*, 103–110.

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*, 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*, 103–131. https://doi.org/10.1111/lang.12029

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 against grammar correction in L2 writing classes": A response to Ferris. *Journal of Second Language Writing*, *8*, 111-122.

Truscott, J. (2001). Selecting errors for selective error correction. *Concentric: Studies in English Literature and Linguistics, 27,* 225-240.

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

Truscott, J. (2016). The effectiveness of error correction: Why do meta-analytic reviews produce such different answers? In Y-N. Leung (Ed.), *Epoch making in English teaching and learning: Evolution, innovation, revolution* (pp. 129-141). Crane.

Truscott, J. (2020). *The efficacy of written corrective feedback: A critique of a meta-analysis.* Unpublished manuscript, National Tsing Hua University, Taiwan. Available on Research Gate.

Truscott, J. (2022). *Efficacy of written corrective feedback in writing instruction: A critique of another meta-analysis*. Unpublished manuscript, National Tsing Hua University, Taiwan. Available on Research Gate.

Truscott, J., & Hsu, A. Y.-p. (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. (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. 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