

The Effectiveness of Error Correction: Why Do Meta-analytic Reviews Produce Such Different Answers?

John Truscott
National Tsing Hua University
truscott@mx.nthu.edu.tw

Meta-analysis provides a useful tool for reviewing and synthesizing the research on a given topic and, hopefully, answering important questions about that topic. Applied to error correction, though, meta-analytic reviews have produced a seemingly bewildering variety of conclusions, ranging from strong support for the effectiveness of correction to strong rejection of the practice as ineffective and possibly harmful. This paper argues that positive conclusions are reached only when the meta-analysis does not connect to genuine teaching concerns. The most important aspect of the problem is the frequent reliance on vague or abstract notions of “effectiveness of corrective feedback”, where “effectiveness” is not defined in terms of any teaching goals. The other, related aspect is insufficient concern with how the research is or can be related to actual teaching practice. When attention is focused specifically on research that examines, under realistic teaching conditions, the effectiveness of correction for the central concerns of language teaching, the conclusion is that research has found correction ineffective.

keywords: error correction, meta-analysis, central pedagogical question

INTRODUCTION

For some time now the effectiveness of error correction has been a topic of discussion and debate (e.g. Ferris, 1999, 2003, 2004; Truscott, 1996, 1999, 2005, 2007a). This debate has important implications for teaching practice, specifically for teachers’ decisions on whether correction is a worthwhile practice. In recent years, many authors have sought to determine the effectiveness of correction quantitatively, combining the findings of many different studies through meta-analytic review. Plonsky and Brown (2015) identified 18 papers in which this was done, their most striking feature being the diversity of the conclusions they reached, ranging from Russell and Spada’s (2006) conclusion that correction has a large positive effect to my own that it has a very small negative effect (Truscott, 2007a).

In this paper I want to suggest an explanation for these contrasts: They occur largely because most meta-analytic reviews have failed, to varying degrees and in varying ways, to address any genuine pedagogical concerns. Most importantly, they do not address what I will call the *central (pedagogical) question*: Does correction help learners develop their ability to use the language correctly *in practical ways*? This problem overlaps with another: the frequent reliance on studies that are far removed from any realistic teaching conditions. Because of such problems, the conclusions of these reviews typically have little to offer in regard to the concerns of language teachers.

In the first section, I will briefly describe meta-analysis. I will then consider the issue of what a meta-analysis on the topic of correction should be trying to accomplish, i.e. what question(s) it should be addressing. The following section then surveys the meta-analytic reviews that have been done on this topic, focusing on the problems just noted. After a brief discussion of bias in meta-analysis, the final section uses this survey to offer an answer to the question of how effective correction is in regard to the central pedagogical question – the answer being that it is not effective.

META-ANALYSIS

Meta-analysis (see Hedges & Olkin, 1985; Light & Pillemer, 1984; Lipsey & Wilson, 2001; Rosenthal, 1991, 1994) is an extremely valuable tool that has acquired an increasingly important role in the field of second language acquisition (see, especially, Ellis, 2000; Norris & Ortega, 2000; Plonsky & Brown, 2015). Results from a number of different studies dealing with a single issue are translated into a common measure, *effect size*. The various results can then be combined to produce

an *average effect size*, giving an overall estimate of how strong an effect a given treatment has (the treatment being error correction in this case), and to compare results obtained in one set of conditions with those obtained in another set of conditions.

Several effect size measures are available. In second language work, the one most commonly used is Cohen's *d*, representing the difference between two means – those of an experimental group and a comparison group, typically – divided by the pooled standard deviations of the two means. In other words, *d* is the number of standard deviations by which the two means differ. So if a correction group were to obtain scores that were one standard deviation higher than those of a comparison group, the resulting *d* would be 1.00. Cohen (1992) provided some general guidelines for interpreting the size of *d*: small effect = .20–.50; medium effect = .50–.80; large effect = .80 and up. Effect sizes below .20 are sometimes referred to as “negligible”. These guidelines were very general; i.e., Cohen did not try to tailor them to differences among fields. Plonsky and Oswald (2014) recently did this tailoring, offering a reinterpretation of effect sizes specifically for research in second language acquisition. In place of Cohen's lower limits for small (.20), medium (.50), and large (.80), they argued that the appropriate cutoffs for SLA are .40, .70, and 1.00.¹

The 18 meta-analytic reviews on correction identified by Plonsky and Brown (2015) reported overall effect sizes ranging from -.155 (Truscott, 2007a), representing a very small harmful effect of correction, to 1.16 (Russell & Spada, 2006), indicating a large favorable effect. The others generally ranged from small to medium, using Plonsky and Oswald's (2014) guidelines.

WHAT IS THE QUESTION?

Perhaps the most basic requirement for carrying out a meta-analysis is a clear statement of the question that the analysis is intended to answer. If the question is inappropriate or overly loose, the information that comes out of the meta-analysis is likely to be confusing or misleading rather than helpful. In this section I will present my view of what the main question should be in the case of error correction.

The Central Pedagogical Question, and Other Possibilities

In judging what is best practice for teachers in regard to correction, the primary question that we have to ask (though not the only one) is whether correction is effective in improving learners' ability to use the language correctly *in practical ways*. This is what I take to be the primary goal of language teaching and therefore the primary criterion for judging the value of a teaching practice. The purpose of my meta-analysis was therefore to offer an answer to this question.

Practical use is not the only possible goal. In some contexts the teacher's primary concern might be to prepare learners for formal language tests, for example. In other cases, we might want to help future teachers develop the intellectual understanding of the language that will allow them to teach grammar² and answer questions students will ask them about it. Another possibility is that the teacher simply wants to help a student improve a particular piece of writing, perhaps for a school publication, without actually rewriting it for the student. Or maybe we want to determine if learners notice and/or understand the corrections they receive. Each of these cases raises its own questions, none of which I will address here.³ My concern is with what I consider the main issue for language teaching: the effect that our teaching practices, correction in this case, have on learners' ability to use the language in practical ways. Throughout the paper I will treat this, again, as the central question.

¹ These guidelines are for comparisons between experimental group and control group. For gains from pretest to posttest, they suggested .60, 1.00, and 1.40.

² I will avoid the question of whether they *should* teach grammar (see Truscott, 2007b).

³ We might also focus on theoretical issues – how research on correction relates to theories in SLA, an issue that I take up elsewhere.

When a decision has been made on what question is to be pursued in a meta-analysis, this decision serves as a guide in subsequent decisions on which studies to include in the analysis and which to exclude, and then on how to analyze them. Nearly all the differences found in the various meta-analytic reviews can be traced to differing decisions on which studies to include, i.e. to the reviewers' inclusion criteria and the ways they choose to apply those criteria.

What Is Practical Use of the Language?

I have suggested that the central question for a meta-analysis is how correction affects practical use of the target language. This inevitably raises the question of what constitutes practical use. There is of course no precise definition. Fortunately, though, it is not difficult to judge if a particular case qualifies or does not qualify or, in some instances, falls between. If learners are using the language to express their ideas or feelings, this is practical use. If they are using it to practice a particular formal point they have been taught, or to demonstrate their mastery of that point, this is not practical use. Examples of impractical use include tests that ask learners to fill in the blank with the proper verb ending, make a sentence that follows the grammar rule they were just taught, judge whether a sentence obeys a grammar rule, construct a sentence in accordance with such grammar rules, or translate example sentences between target language and native language. Results of such tasks do not help us answer the central question. A focus on practical use also requires that the tests avoid any focus on a language form that is being taught, as it is very unlikely that learners will be able to maintain such a focus in normal communicative contexts, especially when they have been taught a very large number of grammar rules. In the great majority of studies found in the literature, it is not difficult to judge where their measures stand on this practical-use criterion (see Truscott, 2005, 2007b).

From a general testing perspective, what I am talking about here is simply validity. If our goal is to measure how well our teaching practices are helping learners achieve the primary goals of language learning, then the measures we use are valid if and only if they succeed in measuring learners' success in achieving those goals. The limits of defining practical use are simply the standard limits of defining valid measures.

To return to a point considered above, this is not to suggest that the central pedagogical question is the only legitimate question that can be pursued in a meta-analysis of correction. Authors who are interested in how correction affects other goals can and perhaps should do meta-analytic reviews of studies that address those questions. The crucial point is that the goals, and the research questions that follow from them, should be explicitly identified. It is essential that we avoid confusion among goals; we must not mix studies that address one question with studies that address a different question.⁴

META-ANALYSES USUALLY DON'T ADDRESS THE CENTRAL QUESTION (OR ANY OTHER PEDAGOGICAL ISSUES)

Given that the primary goal of language instruction is to help learners develop practical ability to use the language, we might expect meta-analytic reviews on correction to be designed and executed in a way that will directly address its contribution to achievement of this goal. This would mean focusing on those studies that directly address the issue and excluding those that do not, or else treating them as a separate and secondary category. But most meta-analyses of correction do not do this. They rely instead on vague and/or abstract notions of "effects" or "effectiveness", with little concern for exactly what these notions mean in the context of actual teaching. The result is frequent inclusion of studies that do not help us answer the central pedagogical question, and exclusion of others that do. Questions like "Is corrective feedback effective?" are not interesting

⁴ Compare Norris and Ortega's (2000, p. 498) caution that researchers should use tests that can support the conclusions they want to draw.

unless there is a clear understanding of what it is claimed to be effective for. In this section I will consider various examples of the problem.

Inclusion of Studies that Did Not Measure Learning

The extreme example of the inclusion problem – one that the majority of reviews have fortunately avoided – is the inclusion of what I have called revision studies (Ashwell, 2000; Fathman & Whalley, 1990; Ferris & Roberts, 2001). In these studies, learners wrote one assignment, received correction on it, and then rewrote it on the basis of the corrections. These results were compared to those of learners who did the same tasks but without any correction. This provides a useful measure of learners' ability to use corrections for editing the assignment on which the corrections are marked, but there is nothing in these studies to tell us whether students learned from this experience. Truscott and Hsu (2008) provided empirical evidence that these findings have no relation to learning, finding that a group which showed the same benefits found in previous revision studies showed no benefits when given a new writing task, performed without the teacher's assistance.

Lee (1997) is a marginal member of this group of studies. Lee asked learners to identify errors implanted in a newspaper article and then to classify and correct them. For one group, all the errors were underlined. Two other groups did the task without the underlining, one getting information about whether each line did or did not contain an error, the other getting no help at all. Not surprisingly, the underline group did better than the other two. Their success, like that in the basic revision studies, is not a measure of learning.

If the goal (or a goal) of a meta-analysis is to determine whether correction can help learners revise the work on which they receive the corrections, then it should deal specifically with these studies. If it is concerned with this question as one part of a more general exploration of pedagogical issues involving correction, then it should provide separate analyses, each explicitly tied to a distinct research question. Findings from revision studies should never be averaged in with findings from studies that measured learning, as the two are addressing fundamentally different questions. Most authors have recognized this point and excluded the revision studies, the main exception being Russell and Spada (2006), a meta-analysis which deserves special attention because it reported the highest overall effect size of all the 18 that Plonsky and Brown (2015) identified.

Inclusion of Studies that Addressed Secondary Questions about Correction and Learning

The more common problem in meta-analyses of correction is the inclusion of studies in which the methodology made them far removed from actual teaching. More importantly, the measures that were used were far removed from practical use of the language. If the goal, again, is to draw conclusions about what works in language classes and to give advice to L2 instructors, the inclusion of such studies in a meta-analysis is counterproductive.

Perhaps the most important meta-analysis dealing with oral correction is Li (2010). One of the strengths of her review is that in addition to providing an overall analysis she separated the studies in terms of a number of potentially relevant variables. This secondary analysis showed that a large percentage of the studies were conducted in laboratories and relied on measures of learning that had little relation to realistic use of the target language. The overall effect size she reported thus represented the combination of these artificial studies with research that did address the central question, with the implication that this number has little to offer to the language teacher. On the other hand, the careful separation and analysis of studies by types provides valuable material for evaluation of how correction fares on the central pedagogical question. I will present such an evaluation below.

If Li (2010) is the most important meta-analysis of research on oral correction, its counterpart for written correction is probably Kang and Han (2015), which was not included in Plonsky and Brown's (2015) survey, presumably because of its recency. Like Li, these authors

included the findings of many studies that were carried out in highly artificial manners with testing that did not reflect the learners' ability to use what they had learned in any practical ways. Like Li, they offered as their primary finding an average of the findings from these studies and from those that did address the central pedagogical question. And, like Li, they presented enough information to allow a judgment of how effective correction is in regard to the central question, which I will consider below.

Similar issues arise in Kao and Wible's (2011) meta-analysis on written correction. The authors' goal was to distinguish the effects of focused correction from those of unfocused correction, and their conclusion was that unfocused is ineffective but focused has substantial effects, with an overall effect size of .762 on immediate tests and .800 on delayed tests. While the effort to separate the effects of different types of correction is commendable, this particular effort raises, in a big way, the issue of what question the meta-analysis is actually addressing. The studies included in the focused category (Bitchener, 2008; Bitchener & Knoch, 2008, 2010a,b; Ellis, Sheen, Murakami & Takashima, 2008; Sheen, 2007; Sheen, Wright, & Moldawa, 2009), with the relatively high effect sizes, did not use the kinds of writing tasks that we would normally assign in our classes but rather special exercises designed specifically for the particular grammar point they were studying. In almost all the studies the testing used this same exercise, further limiting the relevance of the findings for actual language learning and probably raising the scores as well, as a match between training tasks/ conditions and testing tasks/ conditions is beneficial for recall. It is worth noting, in this context, that the study in this group that used more valid testing, Ellis et al., yielded an effect size near the lower limit of Plonsky and Oswald's (2014) "small" range. The authors of these studies also chose, deliberately, to use a very simple grammar point, greatly limiting the generalizability of their findings. Work of this sort does not address the central question; nor does a meta-analysis of such work.

One author who can be credited with some recognition of these issues is Miller (2003; Miller & Pan, 2012). In the abstract of his dissertation, Miller (2003) stated that "The purpose of this meta-analysis is to examine whether oral corrective feedback is effective in the noticing of language learners' errors", avoiding the danger of confusing this awareness of errors with actual acquisition. On the other hand, in Miller and Pan's (2012) study of recasts, they stated (p. 50) that their research questions were "Do recasts have an effect on SLA? If so, which characteristics of recasts have an effect on SLA?" The use of "SLA" is an example of the vagueness problem, and the differences among studies did not play a role in the analysis, but in the discussion (p. 56) they did note an important limitation of their findings, that the overall effect size might greatly overestimate acquisition: "The positive effect sizes seen in the studies included in this analysis may only be an immediate response to the treatment (i.e., uptake...)"

Exclusion of Studies that Directly Addressed the Central Question

The problem of "What is the question?" comes up not only in the doubtful inclusion of some studies, but also in the *exclusion* of others that clearly do address the central issue. I will briefly consider a few cases.

A striking example is the meta-analysis of Biber, Nekrasova, and Horn (2011), who obtained a surprisingly large overall effect size, surprising in that it was far above the results of the other meta-analyses that focused on written correction (see Plonsky & Brown, 2015). These authors excluded studies in which "one group received feedback on content, while a second group received feedback on form; or one group received direct correction of errors, while only general comments were provided to a second group" (p. 25). In other words they excluded all the studies that compared correction to the alternatives used by non-correcting teachers. The logic was that these studies did not "isolate the influence of individual factors". If the goal of doing the meta-analysis is to investigate "pure" correction, abstracted from actual classroom practice, then such a decision may be sensible. If the goal is to provide teachers with useful information for deciding what to do in

their classes, it is counterproductive, as the excluded studies are exactly those that most directly address the central pedagogical question. When this exclusion criterion is taken into account, the surprisingly high overall effect size ceases to be surprising, as the excluded studies typically found correction (very) ineffective.

A similar logic was used by Kang and Han (2015). They excluded the findings of Sheppard (1992) and Polio, Fleck, and Leder (1998) on the grounds that “the effects of feedback could not be isolated from that of other treatments such as conferences” (p. 4). In response to a reviewer’s objection that conferences are not a separate treatment but rather a normal, appropriate part of the feedback process, they stated “Though sympathetic with this view, we, nevertheless, excluded the two studies to ensure greater homogeneity within our dataset, our goal, as it may be recalled, being to investigate the efficacy of L2 written corrective feedback” (p. 13). Again, the realities of language teaching are set aside in the interest of studying an abstract, idealized notion of “corrective feedback”.

Some Other Surprising Decisions on Inclusion and Exclusion

Most of the problems regarding inclusion and exclusion are more or less systematic, based on general principles (principles that often conflict with efforts to address genuine teaching concerns). But a few non-systematic cases should also be noted. One is Kang and Han’s (2015) decision to include Chandler (2003). Their inclusion criteria specified that each included study must have “a control or comparison group (without feedback)”, which was not the case in Chandler’s experiment – her comparison group received error correction. The authors decided to include it anyway “only because Chandler considered it controlled, since students in the comparison group were not required to attend to the feedback they were given until after the study was over” (Note 4, p. 13). Russell and Spada (2006) excluded, without explanation, three studies (Polio et al., 1998; Semke, 1980, 1984; Sheppard, 1992) that are clearly relevant to the central pedagogical question.

Conclusion

I have focused here on the point that meta-analytic reviews typically do not address the central pedagogical question. But the problem can be stated more generally than this. When a meta-analysis relies on vague or abstract ideas of effectiveness and/or is not concerned about how the research is related to actual teaching practice, it does not address *any* genuine pedagogical question.

THE STATE OF THE EVIDENCE

After this discussion of the limits found in most meta-analytic reviews on correction research, the remaining issue is what conclusions we should draw regarding the central pedagogical question: Does correction help learners develop their ability to use the language correctly in practical ways? I will not be concerned with the other questions that might be considered in a meta-analysis of correction research.

The Extremes, and Between

Among the 18 meta-analytic reviews identified by Plonsky and Brown (2015), the most negative conclusion was, again, my own (Truscott, 2007a). My negative findings were the direct result of the decision to include exactly those studies that addressed the central pedagogical question: Does correction help improve learners’ ability to correctly use the language in practical ways? This criterion picked out nine measurements from five studies comparing a correction group and a no-correction group ($d = -.155^5$) and 19 measurements from seven studies that measured absolute gains

⁵ Three analyses were actually done. This average is from the analysis that used one effect size per study. The other two analyses, one including all the measures used in all the studies and the other using one effect size per study except where a study examined two different structures, yielded average effect sizes of $-.204$ and $-.247$, respectively.

made by corrected students ($d = .148$ ⁶). The conclusion was that correction is ineffective and possibly harmful.

This can be contrasted with the meta-analysis that produced the most positive results, Russell and Spada (2006). I noted above that they included three revision studies (Ashwell, 2000; Fathman & Whalley, 1990; Lee, 1997). These studies yielded very impressive effect sizes of 1.98, 1.25, and 2.18, which clearly made a large contribution to the overall result (1.16). The bulk of the remaining 12 studies used teaching and/or testing that made them removed, to varying degrees, from classroom practice and from practical use of the target language. Clear exceptions were Kepner (1991), with a small (or negligible) effect size of .39, and Fazio (2001), for which they mistakenly reported an effect size of .74, the actual number being -.74, as the correction group obtained lower scores than the no-correction group. Excluded from the analysis, without explanation, were three studies that clearly addressed the central question: Polio et al. (1998), Semke (1980, 1984), and Sheppard (1992), with effect sizes of .007, .107, and -.478/-.939 (for verb forms and clause boundaries, respectively). The unusually high overall effect size that Russell and Spada reported, 1.16, was the product of these various decisions on what to include and what not to include (along with the sign error).

This meta-analysis, dominated by studies that did not address the central question, produced the most impressive results. The only other that came close was Miller (2003), which explicitly dealt with noticing of errors rather than learning. Biber, Nekrasova, and Horn (2011) obtained a relatively high effect size of .88 for written correction, falling in Plonsky and Oswald's (2014) medium range. But the relevance of this finding for teaching concerns is canceled by the exclusion criteria they used, as described above. Other meta-analytic reviews, which showed the inclusion problem to varying degrees, produced results clearly higher than mine and clearly lower than those of Russell and Spada (2006). It should also be noted that mine was by no means the only meta-analysis to find correction ineffective: Average effect sizes in the "small" range (and below) are common.

Of all the meta-analyses that draw positive conclusions about the effectiveness of correction, the two that probably attract the most attention and get the most respect are Li (2010) for oral correction and Kang and Han (2015) for written. The overall effect size reported by Kang and Han was .54, in the middle of the "small" range according to Plonsky and Oswald's (2014) standards (endorsed by Kang & Han). Li initially obtained an average of .70/.88⁷ for all the measures she included, but decided that the sample included outliers that had to be removed, lowering the numbers to .61/.64. After reanalysis based on the likelihood that relevant findings were missing from the sample, the numbers were further lowered to .56/.53; i.e., this meta-analysis found a "small" effect for correction. Even these authors who hold that correction is effective can offer only very weak support for their conclusions. And the limited support that they can offer rests on studies that did not address the central pedagogical question. This is the topic of the following section.

Amount of Correction and Effect of Correction

Perhaps the most telling feature of the research findings is one that has received little attention but has been clearly observed in research on both written (Kang & Han, 2015) and oral correction (Li, 2010). If correction is effective, as is commonly claimed in this literature, we should expect to find a clear, strong relation between the amount of correction done and the size of the resulting effect. Brief, one-shot treatments should have relatively limited value, while several treatment sessions, over a few months, should have considerable value. And in fact the research does show a clear and strong relation between amount of correction and size of effect – a clear and strong *negative* relation. The more correction is done and the longer the period over which it is sustained, the lower

⁶ Again, this comes from the "one number per study" analysis. For the other analyses, the numbers were .069 and .115.

⁷ There are two numbers because Li used both a Fixed Effects Model and a Random Effects Model. The details need not concern us here.

the observed effect becomes. Given this fact, it is difficult to understand how so many authors see this body of research as evidence that correction is effective. In any case, there is a clear message here that something is very wrong and that some major reconsideration is needed.

What is wrong is not hard to see. A recurring theme throughout the literature, seen in both written and oral contexts, is that (a) short-term studies carried out in artificial manners using artificial testing yield impressive results, while (b) long-term classroom studies looking at realistic use of the language produce very unimpressive results. Studies of the (b) type are the ones that can tell us whether correction is effective in the important sense, but those of the (a) type typically carry as much or more weight in reviews of the research. It is this confounding of study types, and especially test types, which is responsible for the popular belief that research has found correction effective. The core problem, again, is the recurring failure to go beyond “effects of corrective feedback” and focus on the issues that actually matter for teaching. When we focus on the issue that is central for teaching, we find that correction is ineffective.

A NOTE ON BIAS

It will no doubt be clear to readers that I am biased. I wrote this paper to show that the popular positive view of correction is misguided and that the evidence actually points to its ineffectiveness. This rejection of correction is in fact the point of all my work on this topic, as should be clear to anyone who has followed the debate, since I have always been explicit about my views and my goals. And so a word about bias is in order.

We are all biased. One of the clearest findings in psychology is that bias is an inherent feature of perception, memory, and judgment. It is literally impossible for us to perceive, interpret, or remember things without inserting our knowledge, beliefs, and desires into them. Awareness of this fact is built into standard research procedures in psychology, and all other scientific research that involves people.

This point – the universality of bias – naturally applies to meta-analysis, as it does to all other human activities. It applies most strongly to cases in which the reviewers specialize in the area they are reviewing, as is often the case with error correction. There are other researchers whose specialty is meta-analytic review and who therefore might (or might not) be unbiased in regard to pedagogical questions. These researchers are unlikely to draw strong conclusions or to advise teachers on classroom practice. More common, though, are those of us whose interest is primarily in pedagogical issues and who see meta-analysis as a tool for studying them. In our work, relevant biases are inevitable and must be recognized; a failure to recognize them can lead to serious misunderstandings of the issues and the evidence. The problem appears in its most disturbing form when authors who have clear, strong views on the subject choose to present their work – and particularly their advice to teachers – as objective scholarship.

Meta-analysis is a very valuable tool, and probably an essential one. But there is always a danger that a given meta-analysis will be taken more seriously than it should be taken. An important lesson of Plonsky and Brown’s (2015) discussion is that we must avoid the idea that a meta-analysis represents the objective truth about the research findings on a given topic. Decisions always have to be made about what studies are relevant, what measures within those studies should be included, how to classify the various measures, and how to interpret the resulting numbers – and these decisions can be made in many different ways. The bias of the reviewer is an inevitable part of the decision-making process. No matter how honest, sincere, and careful we all are, it is a simple fact that our perceptions and judgments will reflect our own beliefs and desires. The best approach for a reviewer, then, is to make these beliefs and desires explicit, allowing readers to make a more informed judgment of the things we tell them.

CONCLUSION

Many different authors have used meta-analysis to judge the effectiveness of correction. Why do we get such different answers? Because we ask different questions. When the question is based on vague or abstract notions of “effectiveness”, divorced from the goals and practices of language teaching, relatively favorable answers result. This is because such approaches lead to the inclusion of studies that obtain positive results but are of doubtful relevance to the concerns of teachers, and to the exclusion of studies that are clearly relevant to those concerns and typically yield negative results. If we focus on those studies that are clearly relevant to the primary concerns of language teaching, the answer that emerges is clear: The research has found correction ineffective. This is not a new or surprising conclusion. It is the one I drew in my original paper on correction (Truscott, 1996) and in all those that followed. Long before that paper, Krashen (1982) made very similar points.

There is nothing inherently wrong with studying the effects of an abstract, idealized notion of correction; the findings might well be of theoretical interest. But we have to recognize that such findings are far removed from teaching issues and cannot legitimately be offered as a basis for decisions on teaching practice. Nor is there anything inherently wrong with different reviewers asking different questions within a given research area. There might well be a number of questions that are worthy of meta-analytic review on the topic. Answers to these questions will of course vary, depending on what exactly the question is. The problem comes when the different questions are not clearly distinguished. If we want to know whether correction is helpful for revision of the work on which it is given, or if it sometimes gives learners intellectual knowledge of the language, or if it helps them pass formal grammar tests, then the answer is probably yes. But in regard to the central pedagogical question – Does correction help learners develop the ability to use the language correctly in practical ways? – meta-analytic reviews offer no reason to think the answer is yes and considerable reason to think it is no.

REFERENCES

- 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. Princeton, NJ: Educational Testing Service.
- Bitchener, J. (2008). Evidence in support of written corrective feedback. *Journal of Second Language Writing, 17*, 102-118.
- Bitchener, J., & Knoch, U. (2008). The value of written corrective feedback for migrant and international students. *Language Teaching Research, 12*, 406-431.
- Bitchener, J., & Knoch, U. (2010a). The contribution of written corrective feedback to language development: A ten month investigation. *Applied Linguistics, 31*, 193-214.
- 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.
- 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.
- Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*, 155–159.
- Ellis, N. (2000). Editor's statement. *Language Learning, 50*(3), xi–xiii.
- 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.
- 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: 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. R. (2003). *Response to student writing: Implications for second language students*. Mahwah, NJ: Erlbaum.
- Ferris, D. (2004). The “grammar correction” debate in L2 writing: Where are we, and where do we go from here? (and what do we do in the meantime?). *Journal of Second Language Writing, 13*, 49–62.
- 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.
- Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando: Academic Press.
- 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. DOI: 10.1111/modl.12189
- Kao, C. W., & Wible, D. (2011). *The distinction between focused and unfocused grammar feedback matters: A meta-analysis*. Second Language Research Forum. Ames, IA, USA, 13-16 October, 2011.
- 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.
- Krashen, S. D. (1982). *Principles and practice in second language acquisition*. New York: Pergamon Press.
- Lee, I. (1997). ESL learners' performance in error correction in writing: Some implications for teaching. *System, 25*, 465–477.

- Li, S. (2010). The effectiveness of corrective feedback in SLA: A meta-analysis. *Language Learning, 60*, 309–365.
- Light, R. L., & Pillemer, D. B. (1984). *Summing up: The science of reviewing research*. Cambridge, MA: Harvard University Press.
- Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage.
- Miller, P. C. (2003). *The effectiveness of corrective feedback: A meta-analysis*. Unpublished Doctoral Dissertation, Purdue University.
- Miller P. C., & Pan, W. (2012). Recasts in the L2 classroom: A meta-analytic review. *International Journal of Educational Research, 56*, 48–59.
- Norris, J. M., & Ortega, L. (2000). Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language Learning, 50*, 417–528.
- Plonsky, L., & Brown, D. (2015). Domain definition and search techniques in meta-analyses of L2 research (Or why 18 metaanalyses of feedback have different results). *Second Language Research, 31*, 267–278.
- Plonsky, L., & Oswald, F. L. (2014). How big is ‘big’? Interpreting effect sizes in L2 research. *Language Learning, 64*, 878–912. DOI: 10.1111/lang.12079
- Polio, C., Fleck, C., & Leder, N. (1998). “If I only had more time:” ESL learners’ changes in linguistic accuracy on essay revisions. *Journal of Second Language Writing, 7*, 43–68.
- Rosenthal, R. (1991). *Meta-analytic procedures for social research* (Rev. ed.). Newbury Park, CA: Sage.
- Rosenthal, R. (1994). Parametric measures of effect size. In H. Cooper & L. Hedges (Eds.), *The handbook of research synthesis* (pp. 231–244). New York: Russell Sage Foundation.
- 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). Amsterdam: Benjamins.
- 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*. Unpublished Doctoral Dissertation, University of Minnesota.
- 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., 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.
- Sheppard, K. (1992). Two feedback types: Do they make a difference? *RELC Journal, 23*, 103–110.
- 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. (2005). The continuing problems of oral grammar correction. *International Journal of Foreign Language Teaching, 1*(2), 17–22. Available at <http://www.tprstories.com/ijflt/IJFLTSpring05.pdf>
- Truscott, J. (2007a). The effect of error correction on learners’ ability to write accurately. *Journal of Second Language Writing, 16*, 255–272.
- Truscott, J. (2007b). Grammar teaching and the evidence: A response to Nassaji and Fotos (2004). *International Journal of Foreign Language Teaching*. Available at <http://www.tprstories.com/ijflt/IJFLTJuly07.pdf>
- Truscott, J., & Hsu, A. Y.-p. (2008). Error correction, revision, and learning. *Journal of Second Language Writing, 17*, 292–305.

Bionote

John Truscott is Professor in the Center for Teacher Education at National Tsing Hua University. His research interests include the Modular Online Growth and Use of Language (MOGUL) framework, related areas of cognitive science and linguistic theory, and the value of error correction and grammar instruction in second language learning.