Critical Feedback
on
Written Corrective Feedback Research

NEOMY STORCH*
University of Melbourne

ABSTRACT

Recently we have witnessed a large growth in research on written corrective feedback (WCF). However, the question posed here is: are researchers and L2 writing teachers now any wiser about the efficacy of WCF? I begin with a summary of early studies and some of their major shortcomings. I then examine more recent studies and conclude that, although many of the shortcomings of earlier research have been largely addressed, research findings are still inconclusive. I argue that currently, in the desire to conduct more robust research, the pendulum has swung too far towards experimental studies. Such studies tend to employ 'one off' treatments, often provided on a very restricted range of errors, and ignore the learners' goals and attitudes to the feedback provided and to improvement in accuracy. I conclude by suggesting directions for a more meaningful and ecological valid research agenda on written corrective feedback.

KEYWORDS:
teacher feedback, written corrective feedback, L2 writing, type of written corrective feedback

RESUMEN

En los últimos años la investigación sobre la corrección de trabajos escritos (*written corrective feedback*) se ha incrementado de forma notable. Sin embargo, la cuestión que se plantea en el presente artículo es la siguiente: ¿ha proporcionado este incremento un mayor conocimiento sobre la utilidad de la corrección? Comienzo el artículo haciendo un resumen de los estudios iniciales en este área y de sus principales limitaciones. Tras esto, examino estudios más recientes y concluyo que, aunque muchas de las limitaciones de la investigación anterior han sido subsanadas, los resultados hasta la fecha son todavía poco concluyentes. Considero que, como resultado del deseo de llevar a cabo una investigación más sólida, en la actualidad se ha tendido en exceso a la realización de estudios de carácter experimental. Estos estudios suelen emplear tratamientos de sesión única, prestan atención normalmente a una gama de errores muy restringida, y tienden a pasar por alto las metas y actitudes de los aprendices con respecto al *feedback* recibido y a la mejora de la corrección de los textos (*accuracy*). Concluyo sugiriendo un conjunto de pasos a seguir para que la futura investigación sobre la corrección de trabajos escritos pueda ser más significativa y tener una mayor validez ecológica.

PALABRAS CLAVE
feedback por parte del profesor, corrección de trabajos escritos, escritura en segundas lenguas, formas de corregir trabajos escritos

_____
Address for correspondence: Neomy Storch, School of Languages and Linguistics, The University of Melbourne, Victoria, 3010, Australia. Tel.:+61 3 83445208   Fax +61 3 83448990 E-mail: neomys@unimelb.edu.au

1. INTRODUCTION

Writing teachers tend to give feedback to second language (L2) writers on a range of issues. However, it is feedback on language use, termed written corrective feedback (WCF), which seems to have attracted the most research attention recently. For example, a quick review of the Journal of Second Language Writing shows that in the past four years (2006-2009) alone 16 articles dealing with WCF given by the teacher were published. An almost equal number (17) of articles was published in the previous 10 years. However, as an L2 writing teacher and researcher, the more I read and contemplate about this topic, the more uneasy I become about the current state of the research (including my own research) on WCF.

In this paper I try to articulate these concerns and provide suggestions for future research. The paper is divided into three major sections. The first section provides a brief overview of early research (mid 1980s to 2003) on WCF. I summarise the criticism leveled at the studies namely in terms of research design flaws and problems of comparability. This overview is very brief as these studies, and their shortcomings, have been discussed at some length by a number of other researchers (Ferris, 2004; Guénette, 2007). The second section deals with more recent research (2005 onwards). I note the focus of this research and examine whether this body of research has succeeded in addressing the shortcomings identified with the earlier research. In the final part, I articulate the lingering concerns I have about the more recent studies, particularly in terms of their design and hence ecological validity. I conclude by suggesting, what I consider to be, research directions that may be more meaningful for L2 writing teachers and researchers.

2. EARLY RESEARCH (1980-2003)

This section reviews 11 published and most often cited studies on WCF (see Table 1 below). These studies focused primarily on whether WCF leads to improved accuracy. Some studies also compared the influence of WCF and content commentaries on students' writing (e.g. Fazio, 2001; Fathman & Whalley, 1990; Kepner, 1991; Semke, 1984: Sheppard, 1992) or the differential effect of different types of WCF (Chandler, 2003; Ferris & Roberts, 2001; Lalande, 1982; Robb, Ross & Shortreed, 1986). The distinction between types of WCF was mainly between direct and indirect WCF. Direct WCF refers to the provision of the correct linguistic form or structure by the teacher (Ferris, 2003), and indirect WCF refers to feedback which simply indicates to the writer that an error has been made, usually via a symbol or an abbreviation (e.g. 'vb' representing an error in the use of verbs).

Table 1 summarises the findings reported by these studies on whether WCF leads to improved grammatical accuracy. Of the 11 studies listed, six show that WCF leads to improved grammatical accuracy. However, of those six, three studies only investigated

revised drafts rather than new pieces of writing (Ashwell, 2000; Fathman & Whalley, 1990; Ferris & Roberts, 2001), and in another study (Lalande, 1982) the results were not statistically significant. Sheppard's (1992) study found that the improvement was significant only on one of the two accuracy measures used (correct use of verbs). Furthermore, Sheppard found that the group that received holistic comments outperformed the group that received WCF not only in terms of grammatical accuracy but also in terms of linguistic complexity. Thus, on the basis of these studies, the evidence in support of WCF is not very strong.

Study	*Improved accuracy?*	*But….Qualifying notes*
Ashwell (2000)	Yes	Investigated only revised texts.
Chandler (2003)	Yes	
Fathman & Whalley (1990)	Yes	Investigated only revised texts.
Fazio (2001)	No	
Ferris & Roberts (2001)	Yes	Investigated only revised texts
Kepner (1991)	No	
Lalande (1982)	Yes	Improvement not statistically significant
Polio et al. (1998)	No	
Robb et al. (1986)	No	Investigated only revised texts
Semke (1984)	No	
Sheppard (1992)	Yes	Improvement on one measure (use of verbs) but not on another measure (sentence boundaries) Group which received content feedback outperformed group which received WCF

Table 1: Early research on WCF

Of the studies that considered the effect of different types of WCF, the reported results are somewhat contradictory. For example, Robb, Ross and Shortreed (1986) reported no differences for type of feedback. Lalande (1982) found that students who received indirect WCF made significantly greater gains than those who received direct corrections; whereas Chandler (2003) reported greater gains in accuracy for students who received direct WCF over those who received one of three forms of indirect WCF. There are also mixed findings about the efficacy of different types of indirect feedback. For example, Ferris and Roberts (2001), who investigated the effects of two types of indirect WCF (underlining versus underling and codes), found no significant differences on accuracy between these two types. However, in Chandler's (2003) study, indirect feedback in the form of underlining led to greater accuracy in the long term than underlining plus codes.

Researchers such as Ferris (2004) and Guénette (2007), among others, attribute the lack of conclusive results in support of WCF or of one type of WCF to poor research design and lack of comparability between the studies (see discussion to follow). Others, (e.g. Bitchener, 2008; Ellis, Sheen, Murakami, & Takashima, 2008; Sheen, 2007) argue that the lack of concrete evidence about discernable gains in accuracy is because the students in these early studies were given feedback on all their errors (with the exception of Fazio, 2001) and that such feedback overwhelmed the learners. These researchers refer to oral corrective feedback studies in SLA (e.g. Doughty & Varela, 1998; Han, 2002; Lyster, 2004) which have reported positive effects of oral corrective feedback as a result of intensively targeting a single linguistic feature.

2.1. Research design flaws

A number of criticisms have been leveled at the early studies on WCF particularly in terms of their research design. The most serious research design flaws include:

2.1.1. The lack of a control group
Most of these early studies lacked a control group; that is, a group that did not receive WCF (e.g. Fazio, 2001; Kepner, 1991; Lalande, 1982; Robb et al., 1986). Chandler's (2003) study claimed to have a control group but this group of learners also received WCF. The only difference was that these learners were simply asked not to attend to the feedback until the end of the semester. Chandler (2009) argued that this group constituted a control group because the students did not correct their errors until after the study was completed nor did they seem to pay attention to the errors (although there is no way of verifying this claim). However, Ferris (1999, 2004) and Truscott (1996, 2004) concur that studies which fail to compare the effects of corrective feedback and no corrective feedback do not provide evidence of the effectiveness of corrective feedback.

2.1.2. No new piece of writing
As mentioned earlier, in a number of these studies (Ashwell, 2000; Fathman & Whalley, 1990; Ferris & Roberts, 2001; Robb et al., 1986), the researchers evaluated students' improvement in accuracy by considering only the revised texts. That is, the learners were not required to produce a new piece of writing. The ability to revise does not provide adequate evidence that the WCF had a long lasting effect beyond the revision stage; that is, that L2 learning has taken place (Guénette, 2007; Truscott, 1999, 2004, 2007).

2.1.3. Inappropriate writing task/task conditions

In a number of studies (e.g. Fazio, 2001; Kepner, 1991; Semke, 1984), students' journals were used to provide WCF (and gauge improvement in accuracy). However, as Ferris (2003), among others, has pointed out, journals are unlikely to motivate students to pay attention to grammatical accuracy. Their purpose is usually to encourage writing fluency. Furthermore, even in studies that used perhaps more appropriate tasks, such as compositions or summaries, the writing was done at home (e.g. Ashwell, 2000; Chandler, 2003; Sheppard, 1992), and thus the time spent on the task and whether any additional assistance was available are difficult to determine with any degree of certainty.

2.1.4. Measuring gains in accuracy

The other shortcoming of the early studies concerns the measures used to assess changes in accuracy resulting from the WCF. As Bruton (2009b) points out, researchers need to consider the type of errors in the initial text, on which the student received WCF, and in subsequently written texts. Specifically, to measure gains (or lack thereof) in accuracy, we can only consider whether the errors in the initial text, and on which the learner received WCF, recur in the new text. Errors in the new text which did not appear in the learner's initial text cannot be included in measures of accuracy. Thus to accurately measure changes in accuracy in response to WCF, researchers would need to trace each type of error which received feedback. This is only feasible if the feedback is confined to a limited range of errors.

Ferris (1999) suggests that feedback may be most effective if it focuses on what she terms 'treatable' errors. Treatable errors (e.g. verb tense and form, subject-verb agreement, article usage) occur in a rule-governed way, and may therefore be more amenable to feedback and self-correction. 'Untreatable' errors, (e.g. word choice errors, missing or unnecessary words), on the other hand, are idiosyncratic and thus less amenable to feedback[1]. Thus a number of researchers (e.g. Bitchener, 2008; Sheen, 2007) suggest that WCF should be focused; that is, directed to one error type (e.g. errors in the use of the past simple tense) or a narrow range of errors and at errors which are deemed treatable.

2.2. Lack of comparability

Ferris (2004) and Guénette (2007) have pointed out that, because the early studies differed so much in terms of key parameters, comparison between these studies is problematic. These parameters include:

2.2.1. Populations

In most of these early studies, with the exception of Fazio (2001), the participants were adult university L2 learners. However, in some studies the learners were studying a second language (e.g. Ferris & Roberts, 2001; Polio et al., 1998) and were thus exposed to the language outside the classroom; in others the learners were in a foreign language (FL) context, where exposure to the L2 may be more limited (e.g. Kepner, 1991; Lalande, 1982). Furthermore, although the L2 learners in most of the studies were described as having intermediate L2 proficiency, the L2 proficiency was not always mentioned (e.g. Robb et al., 1986), and even when mentioned, proficiency measures were not clearly defined (e.g. Ashwell, 2000).

2.2.2. Treatment

The treatments varied between the studies in terms of type and frequency. In some studies feedback was given on both grammar and content (e.g. Ashwell, 2000; Semke, 1984); in other studies it was given only on language use (e.g. Chandler, 2003; Robb et al., 1986). More importantly, in most of these early studies feedback was sustained; that is, it was given on a number of pieces of writing over time (e.g. Chandler, 2003; Fazio, 2001). However, in other studies (e.g. Fathman & Whalley, 1990; Ferris & Roberts, 2001) feedback was provided once, only on one piece of writing. This makes comparison between the studies particularly difficult.

2.2.3. Grammatical accuracy measures

The studies also differed not only in what was counted as an error in accuracy but also in how grammatical accuracy was measured. For example, Kepner (1991) used mean number of errors, and these errors included errors in morphology, vocabulary, and syntax. Lalande (1982) also used mean number of errors but only included errors in grammar and orthography. Other researchers used ratio measures such as error/words (e.g. Ashwell, 2000; Chandler, 2003; Ferris & Roberts, 2001) or a ratio of error free T-units to the total number of T-units (e.g. Polio et al., 1998; Robb et al., 1986).

3. RECENT STUDIES (2005 ONWARDS)

Ferris (2004) concluded her overview of shortcomings in existing research on WCF by calling for more robust studies on the efficacy of WCF, a call that perhaps explains the larger volume of research on this topic since 2005. In this section, I analyse twelve more recent studies in terms of whether they have addressed the criticisms leveled at the earlier studies and whether their results are more conclusive about the efficacy of WCF. These 12 studies (see Table 2 below) were selected because they seemed representative of the research direction in the field.

These studies, like the earlier studies, have two foci: investigating whether WCF leads to improved accuracy over time and whether some forms of WCF are more effective than others. However, rather than investigating whether direct or indirect corrective feedback is more effective, many of the current studies have extended this line of investigation to different forms of direct WCF. For example, research by Bitchener (2008) and colleagues (2005; 2009a, b) as well as Sheen (2007) and colleagues (2009) also investigated whether the provision of metalinguistic explanations is more effective than feedback alone. In addition, the studies by Sheen, Wright, and Moldawa (2009) and Ellis, Sheen, Murakami, and Takashima (2008) also investigated whether feedback that is given only on a limited number of errors (focused) is more effective than unfocused feedback.

3.1. Addressing research design flaws

Each of the 12 studies was analysed for whether it addressed the major design flaws identified earlier. Table 2 summarises the results. The last column in the table indicates whether the measures used to capture gains in accuracy relate to the feedback given. As noted earlier, this is only possible if the WCF is focused and the measure used assesses the correct use of the structures targeted by the feedback. For example, in the Bitchener (2008) study, the WCF targeted only two uses of the English articles. To measure accuracy and subsequent gains (or lack thereof), Bitchener calculated the percentage of correct use of these articles in obligatory contexts.

As Table 2 shows on page 36, most (11) of the current studies had control groups, and some (e.g. Sheen et al., 2009; Van Beuningen et al., 2008) had two types of control groups. For example, in Sheen et al.'s (2009) study one control group received no WCF but self edited on the occasions when data were collected (immediate and delayed tests); the other control group received no WCF and only participated in the pre and final delayed test. The researchers could therefore distinguish between the effects of WCF and of writing practice on gains in accuracy.

All the current studies included a new piece of writing. A range of authentic writing tasks were used (not journals) and these were generally completed under timed conditions (with the exception of Ellis et al., 2008). Where the feedback provided was focused (e.g. Bitchener, 2008; Ellis et al., 2008; Sheen, 2007), this enabled the researchers to use accuracy scores (% of correct usage of the targeted structure) that captured changes in response to the feedback provided. Thus, most of the current studies seem to have successfully addressed the flaws in research design identified in the earlier studies. Perhaps what explains this success is that the majority are experimental or quasi-experimental studies, not classroom-based. The only exception is the Hartshorn et al. (2010) study, and this may explain the lack of a control group in that study.

Study	*Control group*	*New writing*	*Writing task/ conditions*	*WCF related to measure of accuracy*
Bitchener (2008)	Yes	Yes	Picture description (30 min)	Yes (focused WCF)
Bitchener & Knoch (2008)	Yes	Yes	Picture description (30 min)	Yes (focused WCF)
Bitchener & Knoch (2009a)	Yes	Yes	Picture description (30 min)	Yes (focused WCF)
Bitchener & Knoch (2009b)	Yes	Yes	Picture description (30 min)	Yes (focused WCF)
Bitchener et al. (2005)	Yes	Yes	Setter (45 minutes)	Yes (focused WCF)
Ellis et al. (2008)	Yes	Yes	Narratives (based on reading) (In class: untimed)	Yes (focused WCF)
Hartshorn et al. (2010)	No	Yes	Short essays (different topics/genres) (10 min)	No (unfocused WCF)
Sheen (2007)	Yes	Yes	Narrative (based on a reading) (12 min)	Yes (focused WCF)
Sheen et al. (2009)	Yes	Yes	Narrative (based on a reading) (15-20 min)	Yes (focused WCF)
Storch (2009)	Yes	Yes	Data commentary & essay (30 min, 60 min)	No (unfocused WCF)
Truscott & Hsu (2008)	Yes	Yes	Narrative based on pictures (30 min)	No (unfocused WCF)
Van Beuningen, De Jong, & Kuiken (2008)	Yes	Yes	Email explaining a topic using a set of pictures (20 minutes)	No (unfocused WCF)

Table 2: Addressing research design flaws

3.2. Comparability parameters

The 12 studies were also analysed in terms of their key parameters in order to determine whether they can be compared. Table 3 summarises the results of this analysis. WCF treatment was analysed for type (e.g. direct or indirect) and for its duration; that is, whether it was given on a number of writing texts (sustained) or whether it was given only once; that is, on one writing task (one shot design).

Study	*Population*	*Treatment: type and duration*	*Accuracy measures*
Bitchener (2008)	ESL low intermediate Language school (adults) New Zealand	Focused (articles) Direct (+ explanation) One shot	% correct usage in obligatory context
Bitchener & Knoch (2008)	ESL low intermediate Language school, adults New Zealand	Focused (articles) Direct (+ explanation) One shot	% correct usage in obligatory context
Bitchener & Knoch (2009a)	ESL low intermediate English Language dept, University NZ	Focused (articles) Direct (+ explanation) One shot	% correct usage in obligatory context
Bitchener & Knoch (2009b)	ESL low intermediate English Language dept, University NZ	Focused (articles) Direct (+ explanation) One shot	% correct usage in obligatory context
Bitchener et al. (2005)	ESL post intermediate Language school (adults) NZ	Focused (3 structures), direct (+ explanation) Sustained	% correct usage in obligatory context
Ellis et al. (2008)	EFL University, Japan	Focused vs. unfocused Direct Sustained	% correct usage in obligatory context
Hartshorn et al. (2010)	ESL Low to mid advanced English Language Centre, adults, USA	Unfocused Indirect (+ error codes) vs. direct Sustained	EFT/T
Sheen (2007)	ESL Intermediate Community college, USA	Focused (articles) Direct (+ explanation) One shot	% correct usage in obligatory context
Sheen et al. (2009)	ESL Intermediate Pre-academic ESL, USA	Focused vs. unfocused Direct One shot	% correct usage in obligatory context
Storch (2009)	Advanced ESL (IELTS$\geq$ 6.5) University, Australia	Unfocused Direct vs. indirect One shot	EFT/T Error free clauses/clauses (EFC/C)
Truscott & Hsu (2008)	EFL High intermediate University, Taiwan	Unfocused Indirect One shot	Error/words
Van Beuningen et al. (2008)	L2 learners of Dutch High School, Holland	Unfocused Direct vs. indirect One shot	Errors/words

Table 3: Addressing comparability issues

As the table shows, the populations in these studies still vary somewhat. Most are ESL adult L2 learners and of intermediate proficiency. However the definition of intermediate is often quite loose. Terms such 'low intermediate' or 'post intermediate' are not always clearly defined. Furthermore, even if the same descriptor is used (low intermediate or intermediate), it is not clear whether it refers to the same L2 proficiency when used in different contexts (e.g. USA vs. New Zealand). In some studies, the learners' L2 proficiency is not mentioned (e.g. Ellis et al., 2008; Van Beuningen et al., 2008).

The feedback treatment varied, but not as extensively as in the earlier studies. In the majority of the studies reviewed (9), participants were provided with direct feedback and often on specific errors. For example, in studies by Bitchener (2008), Bitchener and Knoch (2009a, b), and Sheen (2007), feedback was provided on two uses of English articles (indefinite 'a' for first mention and definite 'the' for subsequent mentions). Furthermore (and of concern, see later discussion) the treatment in many of the studies was uniform, in the sense that feedback was provided only on one piece of writing (one shot), followed by immediate and delayed post tests.

In terms of accuracy scores, a range of measures was used, governed to some extent by whether the feedback was focused or unfocused. In studies where the feedback was focused (e.g. Bitchener et al., 2005), the researchers used the percentage of correct usage of the targeted structure. In studies where the feedback was unfocused (e.g. Storch 2009; Truscott & Hsu, 2008), different ratio scores (EFT/T, Errors/words) were used rather than frequencies (mean number of errors).

Thus overall, in terms of comparability, the studies still vary somewhat on all the key parameters, but there is a noticeable trend towards greater uniformity in research design. This is perhaps attributable to the fact that a number of these studies were conducted by the same research teams.

3.3. Findings

Table 4 summarises the main findings of these 12 studies in terms of the two research questions. The first question is whether WCF led to improved grammatical accuracy in the short term (on revised texts or immediate post tests) and in the long term (on new texts or delayed post tests). The second question is whether some type of feedback is more effective than others.

Study	Does accuracy improve?	Does type of WCF make a difference?
Bitchener (2008)	Yes: Immediate & delayed tests	Yes: Direct WCF+ written & oral explanations (mini lesson) or Direct only superior to direct + written explanations
Bitchener & Knoch (2008)	Yes: Immediate & delayed tests	No effect for type of direct WCF
Bitchener & Knoch (2009a)	Yes: Immediate & delayed tests	No effect for type of direct WCF
Bitchener & Knoch (2009b)	Yes: Immediate & delayed tests	No effect for type of direct WCF
Bitchener et al. (2005)	Yes: Immediate & delayed tests But: only on 2 of the 3 focused structures	Yes: Direct WCF+ individual conference most effective (but only on past T & articles).
Ellis et al. (2008)	No: Immediate post test Yes: delayed post test	No difference (focused vs. unfocused)
Hartshorn et al. (2010)	Yes: treatment group on post test (new writing)	Yes: dynamic WCF (= sustained, frequent) better than traditional error correction
Sheen (2007)	Yes: immediate & delayed tests	Yes: Direct+ written explanation better than direct only
Sheen et al. (2009)	Yes: immediate test Yes: delayed post test but only for focused WCF	No differences (focused vs. unfocused) on immediate test On delayed test only focused WCF led to gains
Storch (2009)	Yes: immediate & delayed post tests	Mixed findings depending on task type/length (direct vs. indirect)
Truscott & Hsu (2008)	Yes: revised text No: new texts	Not investigated
Van Beuningen, et al. (2008)	Yes: revised texts Yes on delayed post tests but only for direct feedback	Yes on revised version (direct best) Yes on new texts – but only direct

Table 4: Findings of current studies

3.3.1. Does WCF lead to improved accuracy?

Unlike earlier studies, where the majority showed no effect for WCF, Table 4 shows that the majority of studies now provide evidence for a positive and statistically significant effect for WCF. Truscott and Hsu's (2008) study was the only study which reported gains on revised texts but not on new texts. In Van Beuningen et al.'s (2008) study, gains were reported for revised texts following direct and indirect feedback, but on new texts only direct feedback

was found to lead to improved accuracy. It should be noted that in both studies, the WCF was unfocused. The other, somewhat puzzling, findings are reported in the study by Ellis et al. (2008) where gains were found in the long term (on delayed post tests) but not in the short term (immediate post tests). Thus the current studies seem to provide evidence that WCF does lead to improved accuracy. However, it is important to note that most of this evidence is based on improvements made on a limited number of English structures: articles, and past tense.

3.3.2. Does type of WCF make a difference?
In terms of the efficacy of particular types of WCF, the research evidence is still inconclusive. Different types of WCF were investigated in these studies: targeted versus untargeted, direct versus indirect, and different types of direct WCF.

3.3.2.a. Targeted vs. Untargeted
The studies by Sheen et al. (2009) and Ellis et al. (2008) seem to suggest that targeted WCF is more effective than untargeted feedback. However, as mentioned earlier, these studies only considered one grammatical structure: use of referential indefinite and definite articles.

3.3.2.b. Direct vs. indirect
A small number of these studies considered whether direct versus indirect forms of feedback are more effective. In the Van Beuningen et al. (2008) study, both direct and indirect feedback were effective on revised texts, but only direct feedback led to statistically significant gains on new texts. In Storch's (2009) study, direct WCF was found to be more effective on short writing tasks (150-200 words long). However, subsequent analysis of longer writing tasks (250-300 words long) suggested that indirect WCF was more effective. In the shorter tasks, the learners were able to memorise the reformulated text (direct feedback). Ellis et al. (2008) point out that theoretically, the distinction between direct and indirect WCF, is problematic. Indirect WCF assumes that the learner already knows the structure (otherwise they could not self correct in response to the feedback). Therefore indirect feedback can only lead to an increase in control of a linguistic form that has already been partially internalized. It cannot lead to new learning (i.e. learning of new linguistic forms). Given the difficulty in determining whether a structure is new or simply needs further practice to become fully internalized, Ellis et al. conclude that the distinction is not worth investigating.

3.3.2.c. Type of direct feedback
Studies investigating whether direct feedback accompanied by metalinguistic explanations is more effective than direct feedback alone have yielded mixed results. For example, in Sheen's (2007) study, direct WCF together with metalinguistic explanations led to greater gains than direct WCF only. However, Bitchener's (2008) study showed that direct WCF alone was

more effective than direct WCF with written explanations. In subsequent studies with Knoch (2009a, 2009b), the different type of direct WCF (with or without written metalinguistic explanations) yielded no significant differences.

4. LINGERING CONCERNS

Although the current studies are better designed and have yielded some promising results for language teachers (and students) in terms of the efficacy of WCF, there are still a number of lingering concerns I have about these studies. In this section I discuss my main concerns.

4.1. Limited range of structures

As mentioned earlier, many of the studies, particularly those which show evidence supporting WCF have focused on a limited number of linguistic structures: the acquisition of the English article system (Bitchener, 2008; Bitchener & Knoch, 2009a, b; Bitchener et al., 2005; Ellis et al., 2008; Sheen, 2007; Sheen et al., 2009), the simple past tense, and use of prepositions (Bitchener et al., 2005).

It is questionable whether we can draw generalizations about the efficacy of WCF on the basis of evidence on only such a limited range of structures (and only in ESL contexts). Furthermore, researchers who focus only on one structure may find few instances of such structures in their students' writing. Xu (2009) points out that in the Ellis et al. (2008) study, some individuals' texts contained only four instances of the targeted use of the article. This means that these participants received relatively little WCF. Xu (2009) also suggests that focusing on one grammatical structure may encourage the students to consciously monitor their use of that structure in the research writing task. And that it is this overt monitoring that may explain why the experimental group outperforms the control group that received no WCF. However, the delayed post tests in Bitchener and Knoch's (2009b) study, which took place six and ten months after the initial feedback and which showed a continued advantage for those who received direct WCF, counter this argument.

4.2. Length and duration of the feedback treatment

Perhaps one of my biggest concerns is the length and duration of the feedback treatment. It is interesting to note that in rebutting one of the most trenchant critics of WCF (Truscott), Ferris (1999:5) argued that Truscott (1996) only referred to studies which "consisted of a 'one-shot' experimental treatment". Yet most (9) of the recent studies, as noted above, used 'one shot' designs, with feedback provided only on one occasion and on a single text (e.g. Bitchener, 2008; Bitchener & Knoch, 2008, 2009a, b; Sheen, 2007; Storch, 2009; Van Beuningen, et al.,

2008). The limited duration of the feedback treatment was compounded when the feedback provided was direct (with or without written explanations). In these studies (e.g. Bitchener & Knoch, 2008), the participants were given only a few minutes to look over the feedback and then asked to write a new text (immediate post test).

The use of one shot treatments and allowing students limited engagement with the feedback provided are perhaps attributable to the fact that these studies are experimental or quasi-experimental rather than being real classroom studies. Brief treatments may be easier to implement and control when conducting experimental studies, but lack theoretical and pedagogical validity. Theories of SLA (e.g. Gass, 2003; DeKeyser, 2007) suggest that learning requires extensive and sustained meaningful exposure and practice. Sheen et al. (2009) admit that feedback needs to be sustained to be truly effective.

4.3. Affective factors

One of the other major concerns with these experimental and quasi-experimental studies is that they tend to ignore affective factors such as attitudes to the type of feedback provided, the feedback provider, and learners' goals. A growing body of qualitative case study research has attested to the importance of these factors in explaining learner response and uptake of the feedback provided (see Given & Schallert, 2008; Hyland, 2003; Hyland & Hyland, 2006; Storch & Wigglesworth, 2010a, b).

Bruton (2009b) argues that researchers need to consider learners' motivation to write, to engage with the feedback received, and to revise. The following excerpt (pseudonyms used) was taken from Storch and Wigglesworth (2010a). In this study the learners worked in pairs and were audio recorded as they processed the feedback received. The excerpt shows quite clearly that although the participants had a very low opinion of the feedback provided, the motive guiding their revision was simply compliance with the researcher's presumed desires. The study found that this pair retained very little of the feedback provided in the long term.
Excerpt 1:

> Gus: huh? I don't think this kind of feedback is good, because …
> Jon: yeah
> Gus: people will tend to memorise this
> Jon: yeah this still crap
> Gus: yeah a feedback should not just give away the   answer. Yeah that's… that's my opinion. Okay so, are we supposed to memorise this?
> Jon: yeah, you got paragraph one and two, I got paragraph three and four
> (Storch & Wigglesworth, 2010a)

5. CONCLUSION: FUTURE DIRECTIONS

The criticisms leveled at the early research on WCF, particularly in terms of the research design, has led to a proliferation of experimental and quasi-experimental studies. Although such studies may be considered more 'robust' in terms of research design, what they lack is ecological validity. Studies which provide feedback on one type of error and only on one piece of writing and in controlled environments are unlikely to be relevant to language teachers because they do not reflect real classroom conditions. Duff (2006) argues that the more controlled and laboratory like the study, the less generalizable are its findings to natural, non-experimental instructional settings.

It is interesting to note that in the rush to criticize and dismiss the early studies, researchers seem to have ignored some of their strengths: most were conducted in real classrooms, where the classroom teacher provided sustained feedback over the semester and on writing tasks that formed part of the academic program, and where students were required to engage with the feedback (e.g. Ashwell, 2000; Chandler, 2003; Robb et al., 1986; Sheppard, 1992). Some of these studies also elicited students' attitudes to the feedback provided (e.g. Ferris & Roberts, 2001; Semke, 1984).

Thus, what I would like to suggest is that for research on WCF to have more pedagogical relevance for language teachers, it needs to try to incorporate some of these strengths. Thus future research on WCF needs to be conducted in authentic classrooms so that the feedback is given within the context of an instructional program, with ecologically valid writing tasks, and where revision is meaningful for the students because it has a clear purpose (e.g. assessment). Such studies need to be longitudinal to allow for more than one treatment occasion (i.e. sustained). As researchers in the field of L1 writing have argued (e.g. Straub, 2000), for feedback to be effective, it also needs to reflect and reinforce what is taught and emphasized in the class.

In terms of type of feedback, Ellis et al.'s (2008: 368) claim sounds convincing: "A mass of corrections directed at a diverse set of linguistic phenomena … is hardly likely to foster the noticing and cognizing that may be needed for CF to work for acquisition." Providing feedback on a large number of errors may overwhelm the learners, not to mention be extremely time consuming for the teachers. However, what I suggest is that rather than choosing one structure, researchers should select several structures. This selection should be guided by students' needs.

Furthermore, in providing WCF, learners' writing goals and attitude to grammatical accuracy need to be taken into consideration. Hyland and Hyland (2006: 220) note that students are ''historically and sociologically situated active agents who respond to what they see as valuable and useful and to people they regard as engaging and credible''. Case studies by Hyland (2003) have shown that students' motivation and confidence in themselves as

writers may be adversely affected by the feedback they receive. Research by Storch and Wigglesworth (2010 a, b), also using a case study approach, has shown that learners' attitudes towards the feedback affects not only whether and how learners respond to the feedback provided, but ultimately whether there is long term learning.

This call for more qualitative approaches to research on WCF does not imply that there is no merit in experimental research on the topic. Rather, what I argue is that in the desire to conduct more robust research, the pendulum has swung too far towards experimental studies. If the aim is to shed light on the impact of WCF on students' writing, then I would like to propose that future studies need to adopt a more qualitative and ecologically valid research design.

NOTES

[1] However, an investigation by Ferris et al (2000) shows the difficulties inherent in predicting the amenability of error types to correction. Although the study found that learners made substantial progress over a semester in reducing "treatable errors" (e.g. in verb tense and form), it also found little or no progress on other "treatable errors" (e.g., noun endings, use of articles). At the same time, some "untreatable errors" (e.g. lexical errors) showed slight progress.

REFERENCES

Ashwell, T. (2000). Patterns of teacher response to student writing in a multi-draft composition classroom: Is content feedback followed by form feedback the best method? *Journal of Second Language Writing,* 9, 227–57.

Bitchener, J. (2008). Evidence in support of written corrective feedback. *Journal of Second Language Writing,* 17, 102–18.

Bitchener, J. & Knoch. (2008). The value of written corrective feedback for migrant and international students. *Language Teaching Research Journal,* 12, 409–31.

Bitchener, J. & Knoch, U. (2009a). The relative effectiveness of different types of direct written corrective feedback, *System* 37, 322–29.

Bitchener, J. & Knoch, U. (2009b). The contribution of written corrective feedback to language development: A ten month investigation. *Applied Linguistics*, 31, 193-214.

Bitchener, J. Young, S. & Cameron, D. (2005). The effect of different types of corrective feedback on ESL student writing. *Journal of Second Language Writing,* 9, 227–58.

Bruton, A. (2009a). Designing research into the effects of grammar correction in L2 writing: Not so straightforward. *Journal of Second Language Writing*, 18, 136–140.

Bruton, A. (2009b). Improving accuracy is not the only reason for writing, and even if it were… *System*, 37, 600– 613.

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, 267–96.

Chandler, J. (2009). A dialogue. Response to Truscott. *Journal of Second Language Writing*, 18, 57-58.

DeKeyser, R. (2007). Skill acquisition theory. In B. VanPatten & J. Wiliams (Eds.), *Theories in second language acquisition*. Mahwah, NJ: Lawrence Erlbaum, pp. 97-113.

Doughty, C. & Varela, E. (1998). Communicative focus on form. In C. Doughty & J. Williams

(Eds.), *Focus on form in classroom second language acquisition*. Cambridge: Cambridge University Press, pp. 114-38.

Duff, P. (2006). Beyond generalizability. Contextualization, complexity, and credibility in applied linguistics research. In M. Chalhoub-Deville, C. Chapelle & P. Duff (Eds), *Inference and generalizability in Applied Linguistics. Multiple perspectives.* Amsterdam: John Benjamins, pp. 65-96.

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.* Cambridge: Cambridge University Press, pp. 178-190.

Fazio, L. (2001). The effects of corrections and commentaries on journal writing of minority- and majority-language minorities. *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,* l-l1.

Ferris, D. (2003). *Response to student writing: Implications for second language students*. Mahwah, NJ: Lawrence 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. Chaney, S. J. Komura, K. Roberts, B. J. & McKee, S. (2000). *Perspectives, problems, & practices in treating written error.* Colloquium presented at the International TESOL Convention, Vancouver, BC, March.

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.

Gass, S. (2003). Input and interaction. In Doughty, C & Long, M. (Eds.), *The Handbook of second language acquisition.* Oxford, England: Blackwell, pp. 224-255.

Given, L. & Schallert, D (2008). Meeting in the margins: Effects of the teacher-student relationship on revision processes of EFL college students taking a composition course. *Journal of Second Language Writing*, 17, 165-182.

Guénette D. (2007). Is feedback pedagogically correct? Research design issues in studies of feedback on writing. *Journal of Second Language Writing*, 16, 40-53.

Han, Z. (2002). A study of the impact of recasts on tense consistency in L2 output. *TESOL Quarterly*, 36, 543-572.

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 wiring accuracy. *TESOL Quarterly*, 44, 84-109.

Hyland, F. (2003). Focusing on form: student engagement with teacher feedback. *System*, 31, 217-230.

Hyland, K. & Hyland, F. (2006). Feedback on second language students' writing. *Language Teaching,* 39, 83–101.

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,* 7, 305-313.

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

Lyster, R. (2004). Differential effects of prompts and recasts in form-focused instruction. *Studies in Second Language Acquisition*, 25, 39-432.

Polio, C. Fleck, C. & Leder, N. (1998). 'If only I had more time': ESL learners' changes in linguistic accuracy on essay revisions. *Journal of Second Language Writing*, 7, 43–68.

Robb, T. Ross, S. & Shortreed, I. (1986). Salience of feedback on error and its effect on EFL writing quality. *TESOL Quarterly,* 20, 83-95.

Semke, H. (1984). The 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-l10.

Storch, N. (2009).   *Comparing type of feedback and processing mode: Pair versus individual processing of feedback on writing*.   Paper delivered at the Second Language Writing Symposium, Phoenix, Arizona, November.

Storch, N. & Wigglesworth, G. (2010a). Learners' processing, uptake and retention of corrective feedback on writing. Case studies. *Studies in Second Language Acquisition*, 32, 1-32.

Storch, N. & Wigglesworth, G. (2010b) Students' engagement with feedback on writing: The role of learner agency.   In R. Batestone (Ed.), *Sociocognitive perspectives on language use and language learning*. Oxford: Oxford University Press, pp. 166-185.

Straub, R. (2000). *The practice of response: Strategies for commenting on student writing*. Cresskill, NJ: Hampton Press.

Truscott, J. (1996). The case against grammar correction in L2 writing classes. *Language Learning,* 46, 327-369*.*

Truscott, J. (1999). The case for ''the case for grammar correction in L2 writing classes": a response to Ferris. *Journal of Second Language Writing* 8, 111–122.

Truscott, J., (2004). Evidence and conjecture on the effects of correction: a response to Chandler. *Journal of Second Language Writing* 13, 337–343.

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

Truscott, J., & Hsu, A. Y. (2008). Error correction, revision, and learning. *Journal of
Second Language Writing*, 17, 292–305.

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

Xu, C. (2009). Dialogue. Overgeneralization from a narrow focus: A response to Ellis et al. (2008) and Bitchener (2008). *Journal of Second Language Writing*, 18, 270-275.