



# Is feedback pedagogically correct? Research design issues in studies of feedback on writing

Danielle Guénette\*

*Département de linguistique et de didactique des langues, Université du Québec à Montréal, Montreal, Canada*

---

## Abstract

The debate continues between those who believe in giving corrective feedback to students to improve their written accuracy and those who do not. Indeed, the results of the many experimental studies on written corrective feedback carried out over the last 20 years have been so contradictory that second language teachers looking to support their pedagogical choice to correct, or not correct, the grammar of their students' written production are left in the midst of controversy. In this article, I examine these studies from a different perspective. Rather than interpret the conflicting results as a demonstration of the effectiveness or ineffectiveness of corrective feedback on form, I suggest that findings can be attributed to the research design and methodology, as well as to the presence of external variables that were beyond the control and vigilance of the researchers.

© 2007 Elsevier Inc. All rights reserved.

**Keywords:** Corrective feedback on writing; Feedback on form; Second language writing; Error correction; Teacher feedback; Secondary school students

---

Should teachers spend hours correcting their students' written productions? This is a fundamental question for second language teachers who are trying to help their students develop fluency and accuracy in their second language.

As a high school teacher of English as a Second Language (ESL) in the francophone school system of Quebec, I experimented with direct corrections – and then watched my students throw their corrected “written production” into the wastepaper basket before leaving the classroom! I also tried peer corrections and group corrections, but both led to mixed results. I tried indirect feedback, explicit and implicit, and I provided feedback both on form and content. Regardless of

---

\* Tel.: +1 514 987 3000x1907.

E-mail addresses: [guenette.daniele@uqam.ca](mailto:guenette.daniele@uqam.ca), [guenetted@hotmail.com](mailto:guenetted@hotmail.com).

the technique I used, the better students seemed to respond well, while the weaker students constantly needed to be pushed, reminded, and encouraged to write and rewrite. I eventually came to the conclusion that no matter what teachers did, some students would benefit from focused instruction and corrective feedback while others would not. But how can a dedicated teacher believe that to be true and still retain the motivation to go into the classroom and make a difference? This is what triggered my interest in wanting to find out more about corrective feedback strategies.

After reading the many studies on corrective feedback carried out since the early 1980s, I am not much further ahead than I was when I first started teaching. Should teachers provide corrective feedback on form or should they not? The debate still rages between proponents of both options because research so far has not been able to prove, beyond a reasonable doubt, that providing corrective feedback is a decisive factor in the attainment of language fluency and accuracy.

This uncertainty, however, has different consequences whether one is a teacher or a researcher. Researchers will see this as an opportunity to keep looking, refining their questions and hypotheses, their research design, and their procedures along the way. Teachers, on the other hand, who are looking for short or long term results (the end of the semester or the end of the year), feel left out on a limb. What are competent and devoted teachers to do? Should they abandon all forms of corrective feedback because some very well known studies and experts have provided evidence that corrections do not work, or should they continue providing corrections because other studies and experts have demonstrated that under certain conditions, with certain student populations, and in some contexts, error correction is effective? For a teacher–researcher such as myself who has first-hand experience of different types of corrective feedback with hundreds of high school students, and who is now engaged in the pre-service training of second language teachers, a more fruitful avenue would be to understand why and how there can be such disparate results on the efficacy of providing corrective feedback on our students' written work.

In this paper, inspired by the work of Ferris (2004), who states that we are at present unable to confirm that error correction works, as the existing research is “fundamentally incomparable because of inconsistencies in design” (p. 50), I endeavor to further our understanding of the results of studies on written corrective feedback by looking specifically at their research design – the population, the treatments, and the procedures – in the hope that seeing them from this perspective might shed a different light on the findings. As well, there is a “constellation of moderating variables that could make a difference regarding corrective feedback effectiveness” (p. 156) as we are reminded by Russell and Spada (2006) in the concluding remarks of their meta-analysis of research on the effectiveness of corrective feedback for second language acquisition. Therefore, a look at extraneous variables might also be in order. This review will then attempt to answer the two following questions:

- (1) To what extent might the conflicting results of the selected written corrective feedback studies be attributed to the *research design and methodology*?
- (2) To what extent might these results be attributed to variables that were not considered in the design?

In order to answer these questions, I will examine the experimental studies on error correction reviewed by Truscott (1996) and later by Ferris (2004) to see if the lack of positive effects for written corrective feedback shown in many studies might be seen as inconsistencies in the research design, rather than as evidence that feedback does not work. Therefore, although the

findings of the studies will be presented, my focus will be on the research design and methodology. In the first section, I will discuss parameters such as the population, the comparison between groups, and the design itself (longitudinal versus cross-sectional). Next, I will review the treatments (type of feedback provided and how it was provided) and the procedures to show that dissimilarities in methodology make it extremely difficult to compare the results. Lastly, I will look at variables that were either not examined, not specified, or not reported in the studies, and discuss the impact these might have had on the results. To help us identify the “constellation of variables” (Russell & Spada, 2006, p. 156) associated with the development of grammatical accuracy on sentence-level errors in writing, I will look at some descriptive work on the issue of error correction. Although the descriptive studies I will discuss were not specifically concerned with the development of grammatical competence, they have the merit of revealing other dimensions of feedback that were ignored in the experimental studies but that might contribute to explaining the results.

## **Research design**

This section addresses how differences in the design features of the experimental studies under review can make comparisons very difficult. The discussion of the first parameter, the population, will address the issue of proficiency levels and the various measures used to assess those levels. Then, studies that did a comparison of correction versus no correction will be examined. Finally, longitudinal and cross-sectional designs will be compared.

### *Population*

In most studies on corrective feedback, the subjects came from natural groups, meaning that the participants were already enrolled in the classrooms where the experiments took place. However it cannot be assumed that all students were at the same level of proficiency. As Frodesen and Holten (2003) claim, students in the same ESL writing class can vary widely in their command of English grammar, in their familiarity with the structures and vocabulary used in academic writing, and in their background in formal instruction. If we are looking to compare the efficacy of teacher feedback across studies, proficiency levels have to be carefully measured and reported. In the studies analyzed, while steps were usually taken to ensure that no significant differences existed between groups in proficiency level, these steps were perhaps not as rigorous as they needed to be, especially when it came to measuring those levels.

For example, Lalande (1982) reported that a pre-test showed no significant differences in writing abilities in German; however, he failed to explain which criteria were used to arrive at this conclusion. Fathman and Whalley (1990) report that their participants were at “similar proficiency levels and had been placed in classrooms according to holistic ratings of a composition” (p. 181). Although holistic scales are widely used to assess students, they might not be “finely-grained” enough for students with similar proficiency ranges (Polio, Fleck, & Leder, 1998, p. 52). In the case of Ashwell’s (2000) study, subjects were assigned to groups on the basis of an unexplained assessment of formal accuracy of the first draft of an essay. This assessment was carried out by the researcher himself, who was also the teacher, and there is no report of inter-rater reliability, a major weakness of many experimental studies as indicated by Polio (2003). Fazio (2001) did not rank her participants and only took into account whether they were of French L1 or of a minority-language background. Through a pre-test on grammatical knowledge and a diagnostic essay, Ferris and Roberts (2001) found varying levels of fluency and accuracy in

their three groups. They proceeded with the experiment nevertheless, but attributed variance in the performance of their groups to the fact that they were different to begin with. One group consisted of immigrant students, while the other was made of international students who had a more solid grammar base, as was shown by the grammatical pre-test, but perhaps less of a sense of what “sounds” right in a language, according to the authors.

Other studies were more attentive to account for student differences. Semke (1984) used pre-tests and post-tests as well as a multiple-choice cloze test and did an analysis of covariance to adjust for the initial differences found in pre-test scores. Robb, Ross, and Shortreed (1986) also pre-tested their participants, using a cloze test and a narrative composition (calculating error-free T-units (EFTs) and number of words in EFT per total words). Since this last measure showed differences between groups, they used it as a covariate in their subsequent analysis.

Some descriptive research that will be discussed later has shown that the overall proficiency level of the students must be considered before deciding when and how to provide error feedback. Since the students’ proficiency levels in most of the studies mentioned above were either not carefully measured or reported, it is nearly impossible for other researchers to replicate the study and lend reliability and validity to the results. We will therefore never know whether the effects or non-effects of feedback might be attributable to the learners’ proficiency levels, rather than the feedback itself.

#### *Comparison between groups: Correction versus non correction*

One anonymous reviewer of this paper pointed out that there was no virtue in using a no-correction group, that is, a real control group that gets no feedback of any kind, since no one proposes “no feedback” as an option for a writing class. In addition, from a pedagogical perspective, if the study is carried out in a “real” classroom, it would appear almost unethical to single out a “no feedback” group (Ferris, 2004). However, if we want to know if error feedback helps L2 student writers, both Truscott (1996) and Ferris (2004) agree that we must compare students who have received grammar correction with students who have not. As Ferris claims, “If correction is important for learning, then the former students should be better writers, on the average, than the latter” (p. 50).

Ashwell (2000), Fathman and Whalley (1990), and Ferris and Roberts (2001) did a correction/no correction comparison and all three studies showed significant effects for the correction groups. Fathman and Whalley (1990) had three groups that either received feedback on form (FF), feedback on content (FC) or a combination of both, and a control group receiving no feedback. In that study, groups receiving both FF and FC showed gains in formal accuracy. Fathman and Whalley’s results were difficult to dismiss, even for Truscott (1996), whose only argument was that although the study showed that students could produce better compositions with their teacher’s help, it did not mean that they would continue to improve in the future.

With a similar design, Ashwell (2000) obtained similar results: All groups receiving feedback made gains in formal accuracy. Ferris and Roberts (2001) provided evidence of substantial positive effects for feedback groups versus the non-feedback group, although the non-feedback group was more successful than the others in correcting lexical errors (word choice). However, these results can be explained – as the authors themselves point out – by differences in proficiency levels between their feedback and non-feedback groups.

As can be seen from Ferris and Roberts’s (2001) results, having a control group is not sufficient. What is needed is a control group that is in every way comparable to the experimental groups in terms of proficiency level, writing conditions, and instructional context. Only then can

the correction/no correction comparison be really informative and help further our understanding of the effectiveness of error correction. Although most teachers would probably prefer to believe the assumption that feedback works, evidence that students receiving no error treatment correct their errors as often or effectively as students receiving treatment, for example, might bring teachers to develop other pedagogical techniques to address the grammatical needs of their learners. On the other hand, finding the opposite would confirm that the time spent correcting the students' errors is not in vain. Comparing correction and non-correction groups would also provide evidence to confirm or disconfirm Truscott's (1996) claim that a non-correction group might perform as well, better, or worse than a correction group (Truscott was referring particularly to Lalande (1982) and Robb et al. (1986) who did not have a "no-correction" group).

### *Longitudinal and cross-sectional designs*

The three studies mentioned above that found significant effects for students receiving feedback over students who were in the no correction group were of short duration (Ashwell, 2000; Fathman & Whalley, 1990; Ferris & Roberts, 2001). Fathman and Whalley's participants wrote one essay and had 30 minutes to correct it, while Ashwell's (2000) students wrote three drafts of one composition. In Ferris and Roberts's (2001) experiment, students wrote an essay which was corrected and given back two weeks later and which they had twenty minutes to correct. In all cases, results showed that students improved their accuracy on a particular piece of writing if they were guided by their teacher. We cannot conclude from these studies that feedback has lasting effects on improvements in accuracy; however, the fact that positive results were seen in the short term shows that pedagogical intervention that pushes learners to pay attention to language is useful. But, as one reviewer noted, few or no researchers would deny that students who get corrective feedback on a draft and make the suggested revisions do not improve on a second draft. What is of particular interest to teachers is whether or not students can sustain this improvement on subsequent writing. This is why we need more longitudinal studies such as the ones discussed below that have traced the development of accuracy over time.

Chandler (2003), Fazio (2001), Goring-Kepner (1990) Polio et al. (1998), Semke (1984), and Sheppard (1992) examined whether feedback could help students improve their writing accuracy, and whether positive effects would stand the test of time. Polio et al. (1998) used a longitudinal design, but they also looked at how students could improve a particular piece of writing if they were given time and guided help. The pre- and post-tests consisted of a 30-minute essay with a 60-minute revision and there was a 15-week lapse between the two, during which students wrote journal entries. The experimental group was given additional grammar exercises and asked to revise one of two journal entries. This group improved both over the semester and from draft to revision, but did not perform any better than the control group on measures of linguistic accuracy.

Goring-Kepner (1990) experimented with feedback on form and feedback on content, and her participants were given feedback six times before their production was analyzed for grammatical accuracy. Results indicated that students who had been corrected on form were not any more accurate than students who had received feedback on content. Fazio (2001) provided weekly feedback to pupils randomly assigned to feedback on form, feedback on content, or a combination of both. Her experiment lasted five months and the pupils (primary level children) were regularly reminded by their instructor to review the comments and corrections before engaging in additional writing. Results showed that, regardless of the feedback provided, over time, the subjects increased the number of errors they committed in grammatical spelling (noun/adjective agreement and subject/verb agreement in French). Semke (1984) compared feedback

on content with more than one kind of feedback on form over a 10-week semester. She found no differences on measures of accuracy, but the groups that got comments only showed more progress on measures of fluency. Sheppard (1992) also carried out his experiment over a 10-week period and his findings resemble those of Semke and Goring-Kepner, in that there was no positive effect for error feedback. In addition, there was a significant advantage for grammatical accuracy on punctuation for the content-only group.

Chandler (2003) also did a semester-long experiment in which students wrote five essays that were collected every second week. The experimental group corrected the errors that had been underlined before submitting their next essay, while the “control” group corrected their errors only at the end. Chandler found that the accuracy of students who were required to correct their errors before submitting their next assignments improved over the semester, but that both groups increased on a measure of fluency.

Studies that examined whether feedback on form could help improve students’ writing in the short-term showed positive results (Ashwell, 2000; Fathman & Whalley, 1990; Ferris & Roberts, 2001) while studies that examined whether feedback could help students improve their accuracy over time did not show significant effects for feedback (Fazio, 2001; Goring-Kepner, 1991; Polio et al., 1998; Semke, 1984, Sheppard, 1992). The only study that found significant positive effects for feedback on accuracy was Chandler’s (2003), but it does not offer solid evidence in favor of feedback since both the control and experimental groups saw their errors indicated, the only difference being what they were asked to do with the feedback (i.e., correct now or correct later).

From these results, we would be tempted to believe that feedback on form does not play a role in the development of accuracy over time. However, a closer look at some of the above-mentioned studies shows that the elicitation task and the feedback techniques chosen might have had a greater effect than was expected. We will return to these issues in the following section, in which we focus on the methodological aspects of the research: the treatment (types of feedback and how it was provided) and the procedures.

## Research methodology

### *Treatment and instruments*

The treatment is the crux of the matter as it is the independent variable that will ensure the study’s external and internal validity and will make it possible for the research findings to be either replicated or generalized. In the studies reviewed below, the treatments varied between combining feedback on form and feedback on content and experimenting with different types of feedback on form. Ashwell (2000), Fathman and Whalley (1990), Fazio (2001), Frantzen (1995), Goring-Kepner (1991) and Sheppard (1992) all included groups who were receiving content feedback or feedback on form, or both. Chandler (2003), Ferris and Roberts (2001), Lalande (1982), and Robb et al. (1986) provided all their students with different types of feedback on form but no feedback on content. In both Polio, Fleck, and Leder’s (1998) and Frantzen (1995)’s studies, the experimental group received error correction and additional grammar exercises or daily grammar review. Semke (1984) provided two types of corrections on form combined with feedback on content (Fig. 1).

In the studies that provided feedback on form only – Chandler (2003), Ferris and Roberts (2001), Lalande (1982), and Robb et al. (1986) – different feedback techniques were used. All studies provided *indirect* corrections (errors are identified but not corrected), of which the various

Feedback on form & feedback on content	Feedback on form only
Ashwell (2000)	Chandler (2003)
Fathman and Whalley (1990)	Ferris and Roberts (2001)
Fazio (2001)	Lalande (1982)
Goring-Kepner (1991)	Robb, Ross and Shortreed (1986)
Semke (1984)*	
Sheppard (1992)	

\* Semke (1984) provided feedback on content and 2 different types of feedback on form.

Fig. 1. Type of feedback provided.

types are listed below, and most studies incorporated direct corrections (i.e., errors were corrected) (Fig. 2).

For comparison purposes, studies with a similar design in terms of treatment will be discussed separately. We will look at studies that compared feedback on content with feedback on form before addressing those studies that experimented with feedback on form only.

*Studies combining feedback on content and feedback on form*

Probably the most often cited research that showed positive results for feedback is the study carried out by Fathman and Whalley (1990). In that study, groups receiving both feedback on form and feedback on content showed gains in formal accuracy. With a similar design, Ashwell (2000) obtained similar results, a fact that speaks to the reliability of Fathman and Whalley’s design: Students who received both form and content feedback were also the ones with the largest gains in formal accuracy. Both studies had a control group that received no feedback of any kind.

At first glance, it would appear that, contrary to Ashwell (2000) and Fathman and Whalley (1990), some longitudinal studies (Fazio, 2001; Goring-Kepner, 1991; Semke, 1984; Sheppard, 1992) provided evidence that corrective feedback does not improve students’ writing over time. The first three studies looked at the development of accuracy through journal writing, an

Indirect corrections
Errors coded
Errors circled
Errors underlined
Errors underlined and coded
Errors underlined + description of error
Errors counted in the margin, but neither marked nor coded

Fig. 2. Different types of indirect corrections.



instrument that is not normally corrected or used for measuring accuracy. Dialogue journals are commonly used in the second language classroom to encourage students to write because they foster a comfortable and positive writing environment. Teachers respond to these journal entries with comments or questions but generally do not grade or correct the students' writing (Wang, 1996). In addition, Ferris (2003) warns us that feedback and correction of journal entries are not likely to have much effect since they are almost never revised by students (p. 126).

In Fazio's study (2001), interview data showed that the pupils paid very little attention to corrective feedback, but generally more to comments. In this case, as in the work of Goring-Kepner (1991) and Semke (1984), the type of instrument arguably had a greater effect than was expected. Students wrote in a journal called *Ma Libert  * ("My Freedom") and these writing moments were their only respite from predominantly form-focused classroom instruction.

In these three studies, therefore, the treatment and the duration of the treatment were valid, but the elicitation task was perhaps not the most appropriate to look specifically at the development of accuracy.

As for the last study in this group (Sheppard, 1992), the results were clearly in favor of responding to content through requests for clarification, as opposed to commenting on form only by indicating the type and location of errors in the margins. In this study, as in others (Ferris & Roberts, 2001; Robb et al., 1986), the participants were foreign language students. Both Ferris (1999) and Leki (2003) remind us that these students might be less motivated to revise and correct their work because they need to write only within the language classroom. Also, as we will see in the next section, simply indicating the type of error and the location in the margin is not necessarily the most effective error correction technique (Chandler, 2003).

As shown in the previous discussion, some studies provided evidence that content feedback was just as effective, if not more, than feedback on form. However, with the exception of Semke (1984), these studies compared only one type of feedback on form with feedback on content and it cannot be assumed that all types of feedback on form are equal. Therefore, studies that compare different types of feedback on form are needed to help us better understand how students at different proficiency levels can benefit from different ways of providing corrective focus on form. A review of these studies follows.

### *Studies of feedback on form*

In this second group of studies, researchers compared different types of feedback on form longitudinally (Chandler, 2003; Lalande, 1982; Robb et al., 1986), except for Ferris and Roberts (2001), whose experiment was a one-time occurrence. Lalande (1982) compared indirect corrections (coded) and direct corrections but asked all students to rewrite or recopy their compositions, and his participants did five in-class essays under the same time and topic constraints. Because there was not a "no feedback" group, the effects of correction could not be compared with the effects of non-correction, but significant effects were found for indirect corrections over direct corrections.

Chandler (2003) examined four different types of feedback and found the opposite. The originality of her design lies in the fact that all students received, in an alternating fashion, four different types of feedback on different pieces of writing. Five essays were collected and corrected every second week and students had to correct their errors before submitting their next assignment. The most significant effects on accuracy were observed with the direct correction and underlining treatment, while the other two treatments – codes and underlining with codes – did not show positive effects. This contradicts Lalande's (1982) findings that



the group that had their errors coded was superior to the direct correction group in terms of accuracy.

Although not mentioned by [Chandler \(2003\)](#), her design seems to be a partial replication of the study conducted by [Robb et al. \(1986\)](#), who tested four different types of feedback on form, including direct corrections. The treatments were not given in an alternating fashion, but the students came from four different groups where classroom activities were identical. The experiment was carried out over a 23-week period, and five compositions were corrected, the last one immediately following a two-week vacation. Results showed no significant differences in measures of accuracy or fluency, since all groups improved their writing regardless of the feedback received.

[Ferris and Roberts \(2001\)](#) tested different types of feedback on form and had a control group receiving no feedback. Their two treatments consisted of having errors coded and underlined or underlined only. Since there was a control group, the effects of correction versus non-correction could be compared, thus addressing one of the weaknesses of [Lalande's \(1982\)](#) study, as pointed out by [Truscott \(1996\)](#). Both experimental groups substantially outperformed the control no-feedback group. However, there were no significant differences between the coded group and the non-coded underlined group, which is similar to what [Chandler \(2003\)](#) found with her students who did not do well either with codes (which she called descriptions) or codes with underlining.

Most studies that compared feedback on content with feedback on form found that the former was just as effective as the latter. In the group of studies that looked at different types of feedback on form only, [Lalande \(1982\)](#) found significant effects for indirect corrections over direct corrections. [Chandler \(2003\)](#) concluded the opposite, while [Robb et al. \(1986\)](#) and [Ferris and Roberts \(2001\)](#) showed that all students improved their writing, regardless of the type of feedback provided. As will be shown in the next section, factors related to the data collection procedures might explain some of the results.

## *Procedures*

### *Classroom activities*

It would be logical to expect that when comparing different types of feedback, every other design parameter must remain constant. For example, in [Sheppard \(1992\)](#), both groups had the same instructor and the activities, courses, and writing topics were similar. However, this was not the case with all studies. In [Lalande's \(1982\)](#) experiment, the rewriting activities in the classroom were entirely different for both groups. The students in the direct corrections group were told to incorporate all corrections into their rewritten version, while the students in the experimental group were involved in guided-learning and problem-solving activities. As with the other group, they had 50 minutes to correct all their errors, but they were encouraged to use their grammar review texts and also to seek the help of their peers or teachers. Yet, [Lalande](#) never addresses the effect that these different pedagogical activities might have had on the students. Had the direct correction students also been given the same instructional treatment, they might have benefited more from the exercise. [Semke's \(1984\)](#) design also had a similar flaw. In that study, two different types of feedback on form were compared: One group saw their errors directly corrected and another group was asked to self-correct (errors were coded) and submit a rewrite one week later. However, the groups were “treated” differently. Because the correction group was asked to rewrite their essay rather than write a new one, students in that group wrote only half as much new material as the other groups. It is therefore very difficult to see the effect of the two different types of feedback on form – direct versus indirect corrections – because of a confounding variable

(quantity of writing) that obscures the issue. In addition, correcting students' errors and asking them to recopy their essay is quite different, cognitively, from only pointing out the errors and asking them to self-correct. In these two studies, the fact that students in both groups were asked to do different activities produced a variable that was not taken into account. Lalande's study shows that to tease apart feedback effects from other factors,<sup>1</sup> we have to ensure that all groups receive the same kind of instruction and are engaged in the same type of activities.

Chandler (2003), who, contrary to Lalande (1982), found significant effects for the direct correction group, hypothesized that the direct correction method might have been superior in her study because it acted as a form of recast by providing a model of positive evidence. We know that recasts do not often lead to learner uptake in oral production (Lyster, 1998), but, in Chandler's case, the rewriting activity made the recast salient because it pushed students to "repeat" it, which might explain why they were able to transfer this knowledge to the next draft. If that were the case, Lalande's (1982) subjects should have been able to do the same. Perhaps a more logical explanation for the different results lies in the design of the study. As students were receiving the different types of feedback in rotation, we could expect that learning would have occurred between treatments, so that students who received the direct correction treatment last probably performed very differently from the students who had received it first. Having experimented with the three other feedback methods, they were no doubt alerted to the fact that they should continue paying attention to their errors, no matter what. In other words, the direct correction method might have proven so effective because it was combined with other feedback treatments. It would be interesting to replicate this study with four different groups working with the same teacher, without alternating the treatments. This would probably be the only way to ascertain whether direct corrections are superior or equivalent.

The lack of significant effects reported by Robb et al. (1986) would seem to indicate that all forms of corrections had almost identical results (although they didn't stand the test of time). But in this study as well, the classroom context is another factor to be considered. Robb et al.'s students were receiving classroom instruction that placed a lot of emphasis on grammatical form and sentence structure. Therefore, all groups were perhaps very attentive to form whether or not they were receiving corrective feedback.

If students in the control and experimental groups are engaged in different classroom activities, such as in the studies discussed above, how can the effects of feedback be isolated? The situation is different if this is the variable being investigated, as in the work of Polio et al. (1998) and Frantzen (1995). In both these studies, the activities were different for the experimental and control groups, but the objective was to examine whether corrective feedback would be more effective if combined with other focus-on-form activities.

### *Student incentive*

Another factor that needs to be carefully controlled for if the study is carried out in intact classrooms is the incentive students receive for participating in the experiment, and grades are usually the greatest incentive a student can receive.

In the Chandler (2003) study, the final draft was graded for fluency and accuracy. The students in Goring-Kepner's study were receiving a grade worth 15% of their total course work for journal writing activities. Frantzen's (1995) "grammar" group (i.e., the group receiving corrective feedback plus additional grammar exercises) was graded on grammar as well as on content.

---

<sup>1</sup> And this may not be possible, as pointed out by one anonymous reviewer.

Participants in [Semke's \(1984\)](#) study were graded according to the number of words written (the “content” group) or on a ratio of mistakes to the number of written words (the “form” groups). One very logical explanation for the verve of the “content” group, then, might be that these students were not worried about losing points, while the other three groups probably tended to write less for fear of making too many mistakes.

In the above-mentioned studies, the fact that participants were graded might have played a role in their motivation to pay special attention to form. One way to minimize the effect of grades would be to follow [Polio et al.'s \(1998\)](#) example since in their study, students were given points for completing their writing assignments, regardless of whether they were in the control or experimental group.

As we can see, the context in which the study is carried out, as well as the pedagogical activities that students are involved in during the experiment, might have an impact on the results. In addition, as will be shown by the findings of the descriptive studies discussed below, variables other than the design parameters and methodology might play a role in how students benefit, or not, from corrective feedback.

### **The contribution of descriptive studies**

Although some of the descriptive work that will be examined here did not investigate specifically the development of grammatical competence, it nevertheless provides a mine of information as to the various dimensions of feedback as a pedagogical tool, from the perspective of both teachers and learners.

The studies described above contribute in giving us a clearer picture of not only what feedback is used for (improving grammatical or rhetorical competence, encouraging students to write, etc.) but also how it is administered by the teachers and how it is perceived by the students. Many studies used a case study approach ([Aljaafreh & Lantolf, 1994](#); [Cohen & Calvacanti, 1990](#); [Cohen & Robbins, 1976](#); [Conrad & Goldstein, 1999](#); [Han, 2001](#); [Hyland & Hyland, 2001](#); [Qi & Lapkin, 2001](#); [Swain & Lapkin, 2002](#)), but some looked at larger groups of subjects in a natural setting ([Blain & Painchaud, 1999](#); [Hedgcock & Lefkowitz, 1994](#)). They also used a variety of procedures and data collection methods. All, except one ([Hedgcock & Lefkowitz, 1994](#)), collected writing samples from their subjects, and the more recent studies conducted since the late 90s incorporate interviews, think-aloud protocols, and transcripts of student-teacher conferences.

The merit of these studies is that they reveal other dimensions of feedback, such as students' ability to engage with feedback, the type of errors that benefit from feedback, the inconsistency of feedback provided by teachers, students' perceptions and preferences, and individual differences. For example, [Cohen and Robbins \(1976\)](#) discovered that feedback was ineffective partly because it was inconsistently provided by teachers. In [Hyland's study \(2003\)](#), among the reasons cited by the students for not correcting their errors, was the fact that they often misunderstood their teachers' comments or suggestions. [Hyland and Hyland \(2001\)](#) found that even when teachers offered praise, it was often perceived by students as a way to soften criticism rather than encouragement to keep trying and writing – which may lead or not, to more fluency and accuracy. Another factor that has been brought to light by descriptive research is the ability (or inability) of students to engage in revision after receiving feedback from the teacher. [Conrad and Goldstein \(1999\)](#) looked at the development of rhetorical competence and found that their advanced students were more successful at revising when it concerned giving examples, facts, or details. [Hyland \(2003\)](#) found that the types of errors as well as the level of proficiency of students were

important variables in their ability to self-correct. Similar results were also reported by Qi and Lapkin (2001). Aljaafreh and Lantolf (1994) showed that students may not be developmentally ready to self-correct and must be trained to become adept at correcting their errors. Han's (2001) longitudinal case study reveals that individual differences are an important variable in whether corrective feedback is successful or not. As a result, Han calls for fine-tuning the feedback to the learner's interlanguage problems in order to directly target the causes of errors. The work of Swain and Lapkin (2002) shows that if a collaborative dialogue develops between learners, they will begin to notice and reflect on the quality of language being used.

As mentioned earlier, even if most studies mentioned above were not concerned with the development of grammatical competence per se, they showed that no matter what the purpose of feedback is (whether to improve form or content, or both), we have to keep in mind that students must understand the feedback and be capable of doing something with it. Teachers also must be consistent with their feedback and adapt it to their students' proficiency level and ability to self-correct. The findings of these studies also remind us that our students expect feedback from their teachers and generally feel that it helps them (Hedgcock & Lefkowitz, 1994; Hyland & Hyland, 2001; Schulz, 1996, 2001).

## Conclusion

What conclusion – if any – can be drawn from the results of the studies discussed? The first, and probably most obvious, is that the studies just reviewed are not necessarily comparable because the design and methodology were not constant. The variables to consider include the following: proficiency level, correction/no-correction comparison, design (longitudinal versus cross-sectional), type of feedback provided and how it was provided, procedures, and elicitation tasks. Therefore, the answer to the first question is that differences in research design and methodology are indeed at the root of the different results obtained.

The second conclusion that can be drawn is that confounding variables make it difficult to isolate, inasmuch as this is possible, the effects of feedback from other factors such as classroom activities and whether or not students were graded on their writing. Also, the overview of the various dimensions of feedback has shown that conflicting results might be attributed to some or all of the extraneous variables gleaned from descriptive studies.

As Russell and Spada (2006) remind us, researchers must “investigate similar variables in a consistent manner” (p. 156) so that they do not end up comparing “apples and oranges (and pears, and grapes, and nectarines . . .)” (Ferris, 2004, p. 52). If we want to make headway in proving the superiority of one feedback type over another (or the superiority of feedback over no-feedback, as Truscott (1996) and Ferris (2004) would remind us), we have to ensure that our groups are in every way comparable, we have to look at effects over time, and we must experiment with types of feedback that are appropriate to the students' proficiency levels and developmental readiness. We also have to control for classroom instruction and ensure that similar data collection procedures are used with both control and experimental groups.

In addition, what is needed are different types of research design that will each provide a missing piece of the puzzle. We need designs that address different issues and control as many variables as possible. We also need descriptive studies that will take the whole context into account, in and out of the classroom.

So what now? Should teachers keep on providing corrective feedback on form to their students? My answer is yes, by all means they should! But they must be aware that there is no “corrective feedback recipe.” The success or failure of corrective feedback will depend on the

classroom context, the type of errors students make, their proficiency level, the type of writing they are asked to do, and a collection of other variables that are as of yet unknown. Teachers must not lose sight of the fact that second language acquisition is slow, gradual, and often arduous, and that corrective feedback is only one of the many factors that contribute to that process.

Finally, one subject that has not been broached in this paper but which was probably a variable in all the studies reviewed is individual learner differences, in particular motivation. Why do we write? We usually write because we have something to say, and someone to say it to. But why do second language students write? Why would they want to improve their accuracy in writing? Some might have a strong instrumental motivation to get into a program of study. As Leki (2003) pointed out, however, even at the college or university level, “L2 writers’ life agendas may or may not ever again include writing in English” (p. 328). For the most part, the hundreds of high school students I had the opportunity to work with in recent years had no such instrumental motivation. They wrote to pass the exam or to please me, but very few were genuinely interested in improving their writing skills, just for the sake of good writing.

That being the case, any type of feedback that does not take the crucial variable of motivation into consideration is perhaps doomed to fail. For students to improve their writing, we assume that, among other things, they have to be provided with appropriate feedback, given at the right time and in the proper context. They have to notice the feedback and be given ample opportunities to apply the corrections. However, when everything is said and done, unfortunately, if the students are not committed to improving their writing skills, they will not improve, no matter what type of corrective feedback is provided.

## Acknowledgments

I wish to thank the editor of *JSLW* and both of the anonymous reviewers for their insightful comments, observations, and suggestions on an earlier version of this paper. I am also grateful to Roy Lyster for his guidance and support.

## References

- Aljaafreh, A., & Lantolf, J. (1994). Negative feedback as regulation and second language learning in the zone of proximal development. *Modern Language Journal*, 78, 465–483.
- 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.
- Blain, S., & Painchaud, G. (1999). L’impact de la rétroaction verbale des pairs sur l’amélioration des compositions des élèves de 5<sup>e</sup> année en immersion française. *The Canadian Modern Language Review*, 56, 73–98.
- 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*, 2, 267–296.
- Cohen, A., & Calvacanti, M. (1990). Feedback on compositions: Teacher and student verbal reports. In B. Kroll (Ed.), *Second language writing: Research insights for the classroom* (pp. 155–177). Cambridge: Cambridge University Press.
- Cohen, A., & Robbins, M. (1976). Toward assessing interlanguage performance: The relationship between selected errors, learners’ characteristics, and learners’ explanations. *Language Learning*, 26, 45–66.
- Conrad, S. M., & Goldstein, L. M. (1999). ESL student revision after teacher-written comments: Text, contexts, and individuals. *Journal of Second Language Writing*, 8, 147–179.
- Fathman, A., & 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. (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. (2003). Responding to writing. In B. Kroll (Ed.), *Exploring second language writing* (pp. 119–140). Cambridge: Cambridge University Press.
- 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.
- Frantzen, D. (1995). The effects of grammar supplementation on written accuracy in an intermediate Spanish content course. *The Modern Language Journal*, 79, 329–344.
- Frodesen, J., & Holten, C. (2003). Grammar and the ESL writing class. In B. Kroll (Ed.), *Exploring the dynamics of second language writing* (pp. 141–161). Cambridge: Cambridge University Press.
- Goring-Kepner, C. (1991). An experiment in the relationship of types of written feedback to the development of second-language writing skills. *The Modern Language Journal*, 75, 305–313.
- Han, Z. H. (2001). Fine-tuning corrective feedback. *Foreign Language Annals*, 34, 582–595.
- Hedgcock, J., & Lefkowitz, N. (1994). Feedback on feedback: Assessing learners’ receptivity to teacher response in L2 composing. *Journal of Second Language Writing*, 3, 141–163.
- Hyland, F., & Hyland, K. (2001). Sugaring the pill: Praise and criticism in written feedback. *Journal of Second language Writing*, 10, 185–212.
- Hyland, F. (2003). Focusing on form: Student engagement with teacher feedback. *System*, 21, 217–230.
- Lalande, J. (1982). Reducing composition errors: An experiment. *The Modern Language Journal*, 66, 140–149.
- Leki, I. (2003). A challenge to second language writing professionals: Is writing overrated? In B. Kroll (Ed.), *Exploring the Dynamics of Second Language Writing* (pp. 315–331). New York: Cambridge University Press.
- Lyster, R. (1998). Recasts, repetitions, and ambiguity in L2 classroom discourse. *Studies in Second Language Acquisition*, 20, 55–58.
- Polio, C. (2003). Second language writing research: What we investigate and how? In B. Kroll (Ed.), *Exploring the dynamics of second language writing* (pp. 35–65). Cambridge: Cambridge University Press.
- 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.
- Qi, D., & Lapkin, S. (2001). Exploring the role of noticing in a three-stage second language writing task. *Journal of Second Language Writing*, 10, 277–303.
- Robb, T., Ross, S., & Shortreed, I. (1986). Salience of feedback on error and its effect on EFL writing quality. *TESOL Quarterly*, 20, 83–93.
- Russell, J., & Spada, N. (2006). The effectiveness of corrective feedback for the acquisition of L2 grammar. A meta-analysis of the research. In J. M. Norris & L. Ortega (Eds.), *Synthesizing research on language learning and teaching* (pp. 133–164). Philadelphia: John Benjamins.
- Semke, H. (1984). The effects of the red pen. *Foreign language Annals*, 17, 195–202.
- Sheppard, K. (1992). Two feedback types: Do they make a difference? *REL C Journal*, 23, 103–110.
- Swain, M., & Lapkin, S. (2002). Talking it through: Two French immersion learners’ response to reformulation. *International Journal of Educational Research*, 37, 285–304.
- Truscott, J. (1996). The case against grammar correction in L2 writing classes. *Language Learning*, 46, 327–369.
- Wang, Y. (1996). E-mail dialogue journaling in an ESL reading and writing classroom. *Proceedings of selected research and development presentations*, National Convention of the Association for Educational Communications and Technology, Indianapolis, IN.

**Danielle Guénette** was a high school ESL teacher and is presently a doctoral candidate at McGill University and *professeure invitée* in the Second Language Teaching Program at UQAM, where she teaches methodology and prepares students for their internships at the primary and secondary levels. Her current research interests are corrective feedback on writing, the use of technology for feedback purposes, and teacher training and development.