THE EFFECT OF DIRECT AND INDIRECT CORRECTIVE FEEDBACK ON L2 LEARNERS' WRITTEN ACCURACY

C.G. van Beuningen, N.H. de Jong & F. Kuiken
Amsterdam Center for Language and Communication,
University of Amsterdam

ABSTRACT

Among scholars there is disagreement on the benefits of corrective feedback on second language learners' written output. While some researchers advocate the usefulness of corrective feedback, Truscott claims that all error correction is unnecessary, ineffective, and even harmful, in that it diverts time and energy away from more productive aspects of writing instruction.

Until now, research outcomes cannot settle this debate since only short-term effectiveness of corrective feedback could be demonstrated. Due to methodological shortcomings, results from studies that investigated long-term effects of error correction on accuracy improvement are inconclusive.

By trying to overcome some of these design related drawbacks (i.e. the lack of a proper control group and time-on task differences between treatment groups), the present study intends to make a contribution to the ongoing error correction debate. The effectiveness of direct and indirect corrective feedback was compared to the effect of two control treatments: a treatment that offered students an extra opportunity to practice their writing skills, and a treatment in which students self-corrected their errors without any available feedback. Results show that corrective feedback can be effective in improving students' accuracy: while short-term effects were found for both direct and indirect corrective feedback, only direct feedback proved to have a significant long-term effect. Neither of the control treatments had a significant effect on students' accuracy.

Introduction

As Swain (1995) argues, it is important for teachers to draw on second language learners' productive skills since producing output not only promotes noticing of linguistic features, but combined with feedback also pushes learners' awareness towards the gaps and problems in their interlanguage (IL). Moreover, the offline character of writing provides learners with more time and opportunity to compare the IL output to the target language (TL) feedback, than online oral production does; when speaking, learners might not (always) be able to make an online IL-TL comparison because of a cognitive overload. In writing on the other hand, learners do have time to compare their output with the provided feedback, and as a result are more likely to notice a gap in their IL. Adams (2003) therefore claims that written production and feedback are of special importance for second language acquisition.

A crucial question is what this feedback should look like. A feedback type commonly used in classrooms is corrective feedback: the marking of a student's error by the teacher. Recently, there has been quite some disagreement in the academic field on the benefits of this kind of feedback on learners' written output.

Truscott, the main opponent of error correction, argues that corrective feedback on second language learners' written output is not only unnecessary and ineffective, but even counterproductive (Truscott, 1996; 1999; 2004; 2007). He bases this claim on two types of arguments. On the one hand Truscott indicates several theoretical problems of error correction. He argues that language teachers – when providing corrective feedback – adopt a "…simplistic view of language learning as essentially the transfer of information from teacher to student" (Truscott, 1996, p. 342) instead of realizing that interlanguage development is a complex and gradual process. Moreover, Truscott regards error correction as ineffective on the basis of practical considerations; he doubts whether teachers are capable of providing feedback adequately and consistently, and if so, he still questions students' ability and willingness to use this feedback effectively. Based on these objections, Truscott (1996) explains that it should be hardly surprising that earlier studies did not convincingly prove the effectiveness of corrective feedback, and concludes that all error correction should be abandoned. Truscott (2004) furthermore argues that until its usefulness has been proven by research findings, corrective feedback can only be considered harmful in that it diverts time and energy away from more constructive activities, such as additional writing practice.

Ferris (1999; 2002; 2004) on the other hand, makes a stand for the use of error correction in writing instruction. In her opinion Truscott's conclusions are premature. She reasons that results from prior research have shown to be inconclusive because of its inadequate methodology, with the main problem that most studies did not include a proper control group. She therefore argues that more, well designed research is necessary before any conclusions can be drawn about the (in)effectiveness of error correction in improving students' future writing (Ferris, 2002).

The present study aims to make a contribution to the ongoing error correction debate. In an experimental set-up with three phases we investigate the effect of corrective feedback on students' accuracy in both revision and subsequent writing tasks. We compare the effectiveness of direct corrective feedback and indirect corrective feedback to two different control treatments.

Our study was conducted at Dutch multilingual schools that adopt a content-based approach to second language instruction. Before elaborating on the current study, we will first take a critical look at earlier research addressing both the questions *if* and *how* error correction should be done in writing instruction.

Empirical background

Investigating the effectiveness of corrective feedback

While a lot of studies make claims about the (in)effectiveness of error correction, most of them actually cannot unambiguously interpret their results, since they lack a proper control group that received little or no correction. Studies that *did* include a control group and investigated the short-term effectiveness of error correction (e.g. Ashwell, 2000; Ferris, 1997; Ferris & Roberts, 2001; Sachs & Polio, 2007), found that participants whose errors were corrected were able to make more accurate revisions than those who did not receive any corrective feedback. In contrast, results from studies investigating the effect of corrective feedback on subsequent writing (e.g. Chandler, 2003; Kepner 1991; Polio et al., 1998; Semke, 1984), are inconclusive. Methodological shortcomings might explain the contradicting findings of these studies, as will be clarified in the following paragraphs.

Semke (1984), who compared the effects of error correction to the effects of content-focused comments, found that error correction had no effect on students' accuracy and a negative effect on their written fluency. It cannot be proven however, that these outcomes are linked to the different treatments incorporated in the study, since students in the content-focused condition had twice as much time to produce new material than students who received corrective feedback. Therefore, it might be the case that these results could actually be explained by the differing amount of writing practice (i.e. time-on-task) between the two treatment groups.

Polio et al.'s (1998) study showed that both students who received corrective feedback and students who did not, were able to improve their accuracy over time. However, students in the experimental condition only produced half as many journal entries as the control group did, because of the editing activities they had to perform. Thus it could well be that the beneficial effects of error correction were leveled-out by the advantages of extra practicing opportunities.

Kepner (1991) did not find any significant differences in error-counts between a group that received error corrections and another group that received message-related comments. However, the flaw in this study is – as observed by Chandler (2003) – that students were not required to do anything with the corrective feedback they received. Hence it remains unclear if students processed the feedback they were presented with.

In a two-phase study Chandler (2003) tried to overcome the methodological shortcomings of the studies described above. In the first phase she addresses "the question of whether error correction can be an effective way to improve accuracy of second language writing" (p. 268). Chandler reports a significant advantage of error correction over a lack of corrective feedback.

However, as Truscott (2004) points out, Chandler's control group is not an actual control group, and therefore she is equally unable to make any judgments on the effectiveness of error correction. The problem is that students in Chandler's control group *did* receive corrective feedback, but were not asked to revise their writing before the end of the semester (and the end of the data collection). What she actually compared was the effect of error correction in combination with revision and error correction without revision. Similar to the studies of Semke (1984) and Polio et al. (1998), time-on-task could be the factor[1] explaining the difference in accuracy gains found between the control group and the experimental group.

Direct and indirect corrective feedback

Whereas only a few studies tested an error correction condition against a true control condition, there are quite some studies that examined the relative effectiveness of varying feedback types, with the dichotomy between direct and indirect corrective feedback receiving the lion's share of researchers' attention. While indirect corrective feedback only consists of an indication of an error (i.e. by underlining the error or providing an error code), direct error correction identifies both the error and the target form.

It has been claimed that students would benefit more from indirect corrective feedback because they have to engage in a more profound form of language processing as they are self-editing their output (e.g. Ferris, 1995). However, this hypothesis could not yet be confirmed since results from studies exploring the relative effectiveness of direct and indirect corrective feedback (e.g. Chandler, 2003; Ferris et al., 2000; Frantzen, 1995; Lalande, 1982, Rob et al., 1986) are inconclusive.

A longitudinal study by Lalande (1982) showed that students who received indirect corrective feedback outperformed students in a direct correction group. Frantzen (1995) and Rob et al. (1986) on the other hand, found that direct and indirect corrective feedback were equally effective. A study by Ferris et al. (2000) revealed yet another pattern; whereas indirect correction proved to be most effective in improving students' accuracy in subsequent writing, students who received direct feedback made the most accurate revisions. Finally, as opposed to Lalande (1982) and Ferris et al. (2000), Chandler (2003) found that direct corrective feedback resulted in the largest accuracy gains, not only in revisions but also in subsequent writing.

Research questions

Earlier research does not provide us with conclusive evidence on the question *if* and *how* corrective feedback should be provided. The current study therefore aims at contributing to the error correction debate by trying to overcome some of the methodological shortcomings of prior studies (i.e. lack of a proper control group and time-on-task differences), and addressing the following research questions:

---

1   In Semke (1984) and Polio et al.'s (1998) studies the fact that less time was allocated to writing practice in the error correction condition than in the control condition, could explain why no positive effects of corrective feedback were found. In Chandler's study it works the other way around: the apparent beneficial effect of error correction could also be explained by other factors, such as time-on-task differences.

1. Does corrective feedback help to improve students' accuracy in writing from an initial task to its revision within content-based second language instruction?
2. Does corrective feedback help to improve students' accuracy in writing from an initial task to a subsequent writing task within content-based second language instruction?
3. If so: What kind of feedback (*direct corrective feedback* vs. *indirect corrective feedback*) on written output will be most effective?

Methodology

Setting and participants

The study was conducted at two Dutch secondary schools with multilingual student populations. Around 80% of the students came from a non-Dutch language background, Arabic and Turkish being the most common L1's[2]. Although most students were born in The Netherlands, many of them only started learning Dutch in school (i.e. at age four).

Both schools adopted a content-based approach to second language instruction. The integration of content and language instruction forms the essence of this approach; language does not only play a central role in language classes, but is also of great importance in classes whose overriding focus is on content (e.g. biology, mathematics, and geography classes). The main aim of this approach is to cater for the special needs of second language and low proficiency learners, who might experience problems understanding and acquiring the content due to the linguistic demands of the input (Hajer & Meestringa, 2004). Since our tasks concern topics in the field of biology, the experiment was conducted during biology classes.

The population in this study consisted of three classes of students (N=62) in their second year of secondary pre-vocational education (or *vmbo-t* in Dutch). Students all were around 14 years of age. Within classes, participants were randomly assigned to four different treatment groups, so that treatment and class did not confound.

Since we applied an experimental set-up, our tasks were designed for experimental purposes only; they were not part of the standard biology curriculum. However, all tasks were administered during class periods. The tasks and topics were introduced and explained by the researcher, and the class teacher was present to maintain order.

Treatments

Four different treatments were included in this study; two experimental treatments: (a) direct corrective feedback (hereafter *Direct*) and (b) indirect corrective feedback (hereafter *Indirect*), and two control treatments: practicing writing (hereafter *Practice*) and (d) revision without feedback (hereafter *Self-Correction*).

---

2   Other L1's within the participant group were (in alphabetic order): Aramaic, Bengali, Berber languages, English, Farsi, Hindi, Kurdish, Spanish, and Urdu.

Experimental treatments: direct and indirect corrective feedback

Students in the Direct and Indirect treatment groups received comprehensive direct or indirect corrective feedback respectively on the texts they produced. All feedback was administered by the first author.

Whereas direct corrective feedback took the form of identifying both the error and the target form (c.f. example 1), indirect corrective feedback only consisted of an indication of the error and its category (c.f. example 2).

We sub-divided the form related errors into nine error categories: word form (e.g. verb tense, singular-plural), word choice, spelling, word order, addition or omission of a word, incomplete sentences, punctuation, and capitalization.

Example 1: direct corrective feedback on form related errors

Je moet ~~het~~ trui niet ~~chemish~~ reinigen.   [You should not dry clean the sweater.]
       _de_   _chemisch_

Example 2: indirect corrective feedback on form related errors

Je moet <u>het</u> trui niet chemish**S** reinigen.   [You should not dry clean the sweater.]
(__ = wrong word, **S** = spelling error)

Control treatments: writing practice and self-correction

Writing practice
Students in the Practice group did not receive any feedback, nor were they invited to execute any revision. Instead, they were presented with two new tasks (i.e. one for each task topic, c.f. section on writing tasks for details) to offer them an extra opportunity to practice their writing skills.

This treatment was included in the design to be able to unambiguously distinguish between effects of error correction and time-on-task effects. Students in the Practice group allocate at least as much time to writing as the students in the error correction groups. Hence, if we would find accuracy gains for either of the error correction treatments but not for the writing practice treatment, we could be sure these gains were brought about by the provision of corrective feedback. In this case, we can confidently rule out for our own findings Truscott's (2004) alternative explanation for accuracy differences between treatment groups found in earlier studies.

Self-correction

Participants in the Self-Correction group were asked to revise their texts without any available feedback, i.e. to self-correct their errors.

We included this treatment to be able to set apart effects of error correction from effects of the revision process as such. It would be plausible that students benefit from having a critical look at their writing products, even without the intervention of a teacher (or researcher in this case).

Experimental set-up

The experiment consisted of three sessions (c.f. table 1). In the first session (T1), a receptive vocabulary test was administered to establish students' overall language proficiency. Furthermore, participants performed the first writing tasks for the two task topics included in the study, i.e. the metamorphosis of insects and instructions on how to do the laundry (c.f. section section on writing tasks for details). Before administering the first writing tasks, the researcher introduced each task (by giving a mini lesson) to ensure all participants had comparable background knowledge on the topic in question. Moreover, the different kinds of errors listed in the experimental treatments section were explained by the researcher. All students were also handed a sheet listing the different error types and an example for each error category. They were told that their writing products would not only be marked with respect to content, but that they would also receive feedback on form related errors, such as those on their hand-out.

One week later (T2) students received feedback and revised their texts accordingly, practiced their writing skills once more (i.e. performed a set of extra tasks), or self-corrected their errors without any available feedback, depending on the treatment group they were assigned to. The first part of the second session was spent on separately instructing each treatment group on what was expected from them: the Direct and Indirect groups were asked to copy their texts revising all errors (i.e. both form and content related errors) the researcher gave feedback on. Students in the Indirect group were furthermore instructed on the meaning and use of the error codes in their texts. The Practice group was presented with two new writing tasks. The researcher first shortly introduced the topics of these tasks. Students in the Self-Correction group were instructed to read over their texts carefully and search for elements in need of revision. Even if no such elements were found, students were asked to copy their texts. All treatment groups were given the same amount of time (i.e. 20 minutes per task) to carry out their assignment.

One week later yet (T3), again having received a short introduction of the tasks' topics, all participants were presented with two new writing tasks. All tasks (including the vocabulary test) as well as the feedback were handwritten.

Table 1: experimental set-up

Vocabulary test	Initial tasks Pretest	Treatment				Subsequent tasks Posttest
T1: week 1	*T1: week 1*	*T2: week 2*				*T3: week 3*
overall language proficiency	text quality (form and content)	Direct	Direct feed-back	Revision	No practice	text quality (form and content)
		Indirect	Indirect feed-back	Revision	No practice	
		Practice	No feed-back	No revision	Additional practice	
		Self-Correction	No feed-back	Revision	No practice	

Writing tasks

Two series of productive writing tasks on biology related topics were used in the experiment: one series on the metamorphosis of different insects and one on laundry instructions and symbols[3]. Each series consisted of three tasks, i.e. an initial task and a subsequent writing task, which were performed by all students at T1 and T3 respectively, and an extra task that was performed at T2 by the Practice group only[4]. All tasks were of a similar type – writing an e-mail to a classmate explaining the task's topic on the basis of a series of pictures. Students were instructed to use at least 100 words for each writing task. Tasks were designed in such a way that the content was unproblematic for all students, since the aim of this study was to elicit the effect of the different treatments on linguistic accuracy only.

The reason for including writing tasks on two different topics is that a learner's language proficiency has proven not to be the only factor influencing a score on a particular task. Research on writing assessment showed that other factors, such as a task's topic, might also contribute to a writer's score (Schoonen, 2005). To control for topic influences and to gain generalizability, participants were presented with two task topics.

Language proficiency pre-test

The instrument we used to obtain an indication of students' overall language proficiency was an adapted version of a receptive vocabulary test, called the Hazenberg & Hulstijn test, originally designed by Hazenberg (1994). The Hazenberg & Hulstijn test was designed to assess test takers' receptive knowledge of a 23,550-lemma list (i.e. the H&H list), which was compounded to define a minimal second language vocabulary for non-native university students (Hazenberg

---

3   (Personal) hygiene matters, such as laundry instructions, are often integrated in Dutch biology methods.
4   The tasks within the metamorphosis series concern the metamorphosis of three different insects: 1) butterflies, 2) wasps, and 3) lady bugs. The tasks within the laundry instructions series concern the instructions for three different garments: 1) a T-shirt, b) a pair of jeans, and c) a woollen sweater.

& Hulstijn, 1996). The original test contains 140 target words, each of which is embedded in a contextually neutral carrier sentence. Students are presented with four options from which they can choose the appropriate meaning. If they do not know a target word at all, test takers can tick a fifth option '*I really don't know*'. The adapted version consisted of 108 multiple-choice items. From the original 140 target words 32 were taken out because of their archaic character.

A vocabulary test was chosen to assess students' general proficiency because earlier research findings suggest that vocabulary knowledge can be used as a predictor of overall language proficiency (e.g. Beglar & Hunt, 1999).

Dependent measures

Students' accuracy was the dependent measure in this study. For all writing performances accuracy was calculated as the *number of form related errors per ten words* (number of form related errors/total number of words x 10).[5]

A ratio measure was used to correct for small individual differences in text length. We used a ten-word ratio instead of the more common hundred-word ratio because texts students produced were relatively short, i.e. around 120 words. For all tasks the same researcher marked the errors in the same way.

Table 2 displays the mean accuracy for every treatment group, itemized per Time (i.e. T1, T2, and T3), and Task Topic (i.e. a: metamorphosis, b: laundry instructions). Table 3 shows the scores on the receptive vocabulary test per treatment group.

Table 2: Accuracy at T1, T2, and T3

Task Topic	Treatment group	Mean[a] and standard deviation T1 (pretest)	Mean[a] and standard deviation T2 (treatment)	Mean[a] and standard deviation T3 (posttest)
a: metamorphosis	Direct (N=15)	1.63 (.76)	0.26 (.32)	1.12 (.74)
	Indirect (N=15)	1.15 (.53)	0.39 (.44)	1.43 (.78)
	Practice (N=13)	1.53 (.69)	1.63 (.55)	1.90 (.59)
	Self-Correction (N=15)	1.21 (.56)	1.12 (.55)	1.32 (.75)
b: laundry instructions	Direct (N=18)	1.55 (.79)	0.29 (.30)	1.11 (.60)
	Indirect (N=17)	1.24 (.53)	0.60 (.47)	1.29 (.43)
	Practice (N=14)	1.67 (.45)	1.87 (.54)	1.76 (.91)
	Self-Correction (N=17)	1.19 (.60)	1.07 (.54)	1.41 (.59)

[a] Mean number of form related errors per 10 words

---

5   Although the effect of error correction on linguistic accuracy was our main interest in this study, students also received feedback on content related issues. We will not report on a content-related measure since error numbers were too small to make statistic analysis feasible. The fact that students committed few content-related errors is explicable; tasks were designed in such a way that the content was unproblematic for all students.

Table 3: Overall language proficiency (score on vocabulary test)

Treatment group	Mean[a] and standard deviation
Direct (N=18)	65.17 (12.95)
Indirect (N=17)	60.35 (13.68)
Practice (N=14)	67.07 (14.33)
Self-Correction (N=17)	65.94 (13.09)

[a] Mean number of correct test items out of the 108 total

Results

Overall

To test for initial accuracy differences between treatment groups, we performed a repeated measures ANOVA with accuracy as the dependent variable, Task Topic (i.e. metamorphosis vs. laundry instructions) as a within subject factor, and Treatment (i.e. Direct feedback, Indirect feedback, Practice and Self-Correction) as a between subjects factor. Results revealed no significant difference between groups concerning the number of form related errors per 10 words that were committed (i.e. accuracy) in the initial phase (T1) of either the metamorphosis or laundry instructions tasks ($F (3, 54) = 2.10$, $p = .112$), nor an effect of Task ($F (1, 54) < 1$, $p = .712$) or an interaction effect between Task and Treatment ($F (3, 54) < 1$, $p = .448$).
A one-way ANOVA was used to test for any initial between-group differences in language proficiency. Results showed that there was no significant difference between treatment groups in the score on the receptive vocabulary test (i.e. language proficiency) ($F (3, 62) < 1$, $p = .506$).
These results show that all treatment groups had a comparable accuracy and proficiency level at the beginning of the data collection. Hence, we can be confident that any differences in error counts found later on in the study, cannot be ascribed to initial differences between treatment groups. Moreover, the task's topic proved not to have any significant influence on students' writing performance in the initial phase of the study.

Accuracy in the revision task (T2)

Accuracy at T2 was analyzed using a repeated measures ANCOVA with accuracy as the dependent variable, Task Topic as a within subject factor, and Treatment as a between subjects factor. Language Proficiency (i.e. the score on the receptive vocabulary test), Accuracy T1a (i.e. the number of form related errors per 10 words committed in the first metamorphosis task)

and Accuracy T1b (i.e. the number of form related errors per 10 words committed in the first laundry instructions task) were incorporated as covariates. We included Accuracy T1a and Accuracy T1b (i.e. accuracy in the initial phase) as covariates to account for effects of individual accuracy differences at T1. A significant main effect was found for Treatment ($F (3, 51) = 57.57$, $p < .001$, partial eta-squared = .77). There was no significant effect of Task Topic ($F (1, 51) = 1.79$, $p = .187$), and none of the interactions turned out be significant (all F-values < 1, except Task Topic*Language Proficiency: $F (1, 51) = 3.35$, $p = .073$, and Topic*Treatment: $F (1, 51) = 1.40$, $p = .253$). While the influence of the Language Proficiency covariate was non-significant ($F (1, 51) = < 1$, $p = .905$), Accuracy T1a ($F (1, 51) = 5.14$, $p = .028$, partial eta-squared = .09) and Accuracy T1b ($F (1, 51) = 22.21$, $p < .001$, partial eta-squared = .30) proved to be significant covariates.

Post-hoc pair wise comparisons (using the Bonferroni adjustment) revealed that students in the Direct group were significantly more accurate (i.e. committed significantly less errors) at T2 than students in either of the other treatment groups (all p-values < .012). The Indirect treatment turned out to be the second most effective treatment, being significantly more effective than both Practicing ($p < .001$) and Self-Correction ($p < .001$). Lastly, students in the Self-Correction group outperformed students in the Practice group ($p = .005$). (C.f. table 4.)

The effectiveness of different treatments could also be investigated by analyzing accuracy progress between T1 and T2. We will do so for the Direct, Indirect, and Self-Correction groups. The Practice group, however, will be excluded from this analysis, since students in this group performed two new writing tasks at T2. It would not be valid to directly compare students' performance on these new writing tasks to the performance on the initial writing tasks, since we could not be sure that the different tasks within one sequence are equally difficult. For the same reason, we will not analyse students' progress between T1 and T3 this way either.

Table 2 shows that students in the Direct, Indirect and Self-Correction group gain accuracy between T1 and T2: the Direct group reduced the number of form related errors they committed per 10 words from 1.63 at T1a to 0.26 at T2a, and from 1.55 at T1b to 0.29 at T2b. The number of errors committed by students in the Indirect group decreased from 1.15 at T1a to 0.39 at T2a, and from 1.24 at T1b to 0.60 at T2b. Error counts for the Self-Correction group show a reduction from 1.21 at T1a to 1.12 at T2a, and from 1.19 at T1b to 1.07 at T2b.

The progress made by the Direct, Indirect, and Self-Correction groups between T1 and T2, was analyzed using a repeated measures ANCOVA with accuracy as the dependent variable, Task Topic and Time as within subject factors, Treatment as a between subjects factor, and Language Proficiency (i.e. the score on the receptive vocabulary test) as a covariate. Significant main effects were found for Time ($F (1, 41) = 16.38$, $p < .001$, partial eta-squared = .29), and Treatment ($F (3, 41) = 3.65$, $p = .035$, partial eta-squared = .15). There was no significant effect of Task Topic ($F (1, 41) = < 1$, $p = .637$). Furthermore, only the interaction between Time*Treatment ($F (3, 41) = 34.23$, $p < .001$, partial eta-squared = .63) turned out to be significant. Language Proficiency proved to be a significant covariate ($F (1, 41) = 10.41$, $p = .002$, partial eta-squared = .20). Paired samples T-tests show that while the accuracy gains between T1 and T2 are significant for the Direct and Indirect group ($p < .001$, using the Bonferroni adjustment), the reduction of the number of errors of the Self-Correction group is

non-significant (p = .092).

From the above we conclude that students in the Direct group were able to make the most accurate revisions. The Indirect treatment turned out to be the second most effective in improving students' accuracy, followed by the Self-Correction method. Students in these three treatment groups reduced the number of errors they committed between T1 and T2, but only the accuracy gains of the Direct and Indirect group showed to be significant. Providing students with an extra opportunity to practice their writing, proved not to have any beneficial effects on their accuracy.

Table 4: post-hoc pair wise comparisons T2 based on estimated marginal means

(J)Treatment / (I)Treatment	Direct	Indirect	Practice	Self-Corr.
Direct		$-.369^a$ $(.113)^b$ p=.012^c d=.34^d	-1.369^a $(.114)^b$ p=.000^c d=2.01^d	$-.954^a$ $(.110)^b$ p=.000^c d=1.14^d
Indirect			-1.000^a $(.121)^b$ p=.000^c d=1.67^d	$-.585^a$ $(.108)^b$ p=.000^c d=.80^d
Practice				$.416^a$ $(.116)^b$ p=.005^c d=.87^d
Self-Corr.				

[a] Mean difference (I-J)

[b] Std. Error

[c] Sign. (Bonferroni adjusted for multiple comparisons)

[d] Effect size (Cohen's d)

Accuracy in the subsequent writing task (T3)

Accuracy at T3 was again analyzed using a repeated measures ANCOVA with accuracy as the dependent variable, Task Topic as a within subject factor, Treatment as a between subjects factor, and Language Proficiency, Accuracy T1a and Accuracy T1b as covariates. Results closely resemble those described above, revealing a significant main effect for Treatment (F (3, 51) = 5.47, p < .005, partial eta-squared = .24), no significant effect of Task Topic (F (1, 51) = < 1, p = .806), nor any significant interaction (all F-values < 1, except Task Topic*Accuracy T1a: F (1, 51) = 1.79, p = .187). The Language Proficiency covariate was again non-significant

(F (1, 51) = < 1, p = .689), while Accuracy T1a (F (1, 51) = 6.06, p = .017, partial eta-squared = .11) and Accuracy T1b (F (1, 51) = 16.23, p < .001, partial eta-squared = .24) proved to be significant covariates.

Post-hoc pair wise comparisons (using the Bonferroni adjustment) revealed that students in the Direct group committed significantly less errors at T3 than students in both the Practice (p = .003) and Self-Correction group (p = .029). Students in the Direct group also outperformed students in the Indirect group, but the difference between these two treatments turned out not to be significant (p = .061). There was no significant difference between the Indirect group, the Practice group, and the Self-Correction group (p = 1). (C.f. table 5.)

Table 5: post-hoc pair wise comparisons T3 based on estimated marginal means

(I)Treatment \ (J)Treatment	Direct	Indirect	Practice	Self-Corr.
Direct		-.408[a] (.153)[b] p=.061[c] d=.38[d]	-.579 [a] (.154)[b] p=.003[c] d=1.00[d]	-.437 [a] (.148)[b] p=.029[c] d=.40[d]
Indirect			-.171[a] (.164)[b] p=1.000[c] d=.62[d]	-.029 [a] (.147)[b] p=1.000[c] d=.01[d]
Practice				.142 [a] (.157)[b] p=1.000[c] d=.61[d]
Self-Corr.				

[a] Mean difference (I-J)

[b] Std. Error

[c] Sign. (Bonferroni adjusted for multiple comparisons)

[d] Effect size (Cohen's d)

Summary and discussion

Results show that all students who had the opportunity to revise their writing products (i.e. students in the Direct, Indirect, and Self-Correction groups) produced fewer errors in their revisions than in their initial texts. Nonetheless, only the accuracy gains made by the groups who received corrective feedback (i.e. Direct and Indirect) turned out to be significant. Moreover, post hoc pair-wise comparisons revealed that both students receiving direct corrective feedback and students whose errors were corrected indirectly, significantly outperformed students who

did not receive any feedback and self-corrected their errors. From this we conclude that it was not the revision process as such which brought about the improved accuracy of the Direct and Indirect group. Thus our first research question, whether or not error correction is an effective means to improve students' accuracy from an initial writing task to its revision, can be answered affirmatively.

The present study aimed at testing not only the short-term but also the long-term effectiveness of corrective feedback. Whereas a short-term effect is defined as the effect from an initial task to its revision, the effect from an initial task to a new writing task is considered a long-term effect (e.g. Ferris, 2002). Adopting these definitions, our results show short-term effects for both direct and indirect corrective feedback. Direct feedback also proved to have a long-term effect on students' accuracy. When we consider the long-term effectiveness of indirect correction however, the picture is not as straightforward. Post-hoc pair wise comparisons did not yield a significant difference between the groups receiving direct and indirect corrective feedback, nor a significant difference between the Indirect, Practice, and Self-Correction treatments. Although clear statistical proof is lacking, two observations lead us to suggest that the Direct treatment was superior to the provision of indirect feedback. Firstly, we would like to point out that although the difference between the Direct and Indirect group is not significant, the results do display a trend (p = .061), and the p-value was submitted to a strict Bonferroni adjustment. Secondly, as we can see in table 2, students in the Direct group improved their writing performance between T1 and T3, while students in the Indirect group actually performed worse on the subsequent writing tasks: the Direct group reduced the number of form related errors they committed per 10 words from 1.63 at T1a to 1.12 at T3a, and from 1.55 at T1b to 1.11 at T3b. The number of errors committed by students receiving indirect feedback on the other hand, increased from 1.15 at T1a to 1.43 at T3a, and from 1.24 at T1b to 1.29 at T3b. Based on the above our answer to the second research question, concerning the effect of error correction on subsequent writing, is that some, but maybe not all types of error correction result in long-term accuracy gains. Moreover, a careful conclusion in relation to our third research question on the relative effectiveness of direct and indirect corrective feedback, is that direct error correction appears to be the most effective treatment for this study's population, resulting in both short-term and long-term accuracy improvement.

The fact that direct error correction seemed to be superior to indirect corrective feedback when considering long-term effectiveness, contradicts the prediction in the literature that students benefit more from indirect corrective feedback because they have to engage in a more profound form of language processing as they are self-editing their output (e.g. Ferris, 1995; Lalande, 1982). It may be that the explanation Chandler (2003) gave for her findings also holds for our results: while students who received direct corrective feedback could instantly internalize the correct form, students who revised their texts based on indirect error correction were unable to do so, since they did not know whether their own hypothesized correction was indeed accurate.

Since we found that the opportunity to practice writing did not yield any significant effect on students' accuracy, this study does not support Truscott's (2004) claim that accuracy gains found in earlier studies (e.g. Chandler, 2003) were due to writing practice (i.e. time-on-task differences between treatment groups) instead of error correction. Moreover, our results

challenge his suggestion that the time spent on dealing with corrections could be allocated more efficiently to alternative activities, such as additional writing practice (Truscott, 2004).

As we mentioned earlier, table 2 shows that whereas direct feedback was effective in improving students' accuracy from an initial to a subsequent writing task, students in the Indirect group committed more errors at T3 than at T1. The same is true for students in the two control groups. We would not claim that this accuracy decrease points at harmful effects of the Indirect, Practice and Self-Correction treatments. Alternatively, a reduction of students' motivation as the study went on could be a possible explanation[6]. While the beneficial effects of the Direct treatment were strong enough to still be observable in spite of this motivational factor, potential effects of the other treatments were not. Support for this interpretation could be found in studies that have shown that students highly value corrective feedback (e.g. Hedgcock & Lefkowitz, 1994; Leki, 1991). It might well be that the absence of this kind of feedback interferes with students' motivation.

Although our findings suggest that error correction can be an effective means of improving students' accuracy in writing, many other questions on issues such as the influence of language proficiency on feedback uptake, or the possible avoidance of complex structures due to error correction, remain to be answered by further research.

In his 2007 article, Truscott states that his case against corrective feedback is actually a case against *grammar correction*. He claims that when error correction proves to be effective, this can only be true for the most correctable error types such as spelling errors, not for grammatical errors (e.g. word order errors). It would therefore be interesting to compare the uptake of feedback concerning different error types[7]. Another point of interest would be the lexical and/ or structural complexity of students' written output before and after error correction. One of Truscott's (2004; 2007) alternative explanations for accuracy gains found in earlier studies is *avoidance*. He claims that the corrected students in Chandler's (2003) study for example, might not have gained accuracy because they benefited from corrective feedback, but because they simplified their writing. Truscott argues that it is the immediate goal of corrective feedback to make learners aware of the errors they committed, and that this awareness creates a motivation for students to avoid the corrected constructions (Truscott, 2007). Further research is necessary to test the avoidance hypothesis. Moreover, as suggested in the literature (e.g. Ferris, 2004; Hyland and Hyland, 2006), it might be the case that indirect corrective feedback is not advantageous to lower proficiency second language learners, since they lack the linguistic competence to self-correct their errors. It would therefore be interesting to investigate the influence of language proficiency on the uptake of the different feedback forms. We will address these three issues in a follow-up study (N=280), in which we will furthermore incorporate a postponed post-test to investigate whether or not the effect of error correction will still prevail three weeks after the moment of treatment.

Although the questions raised above should be attended to in further research, our results clearly show that the use of corrective feedback in content-related classes improves students' written accuracy. Moreover, direct error correction seems to be more effective than indirect error correction for this study's population, since the provision of direct corrective feedback did not only lead to more accurate revisions, but also to more accurate performance on a new writing task.

---

6   Since our tasks were not part of the school curriculum, students' performance was not graded. As a result, students' extrinsic motivation was quite low. Moreover, the experimental design of our study required the tasks within one series to be as comparable as possible. The drawback of this set-up might be that students got bored with writing the same kind of assignments.

7   Although it is not feasible to draw definitive conclusions on the (un)correctability of the different error categories included in this study (since the study was not designed with this kind of analysis in mind), a preliminary analysis of grammatical error categories, does show significant positive effects of grammar correction.

Adams, R. (2003). L2 output, reformulation and noticing: implications for IL development. *Language Teaching Research 7*(3), 347-376.

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*(3), 227-258.

Beglar, D. & Hunt, H. (1999). Revising and validating the 2000 word level and university word level vocabulary tests. *Language Testing 16*(2), 131-162.

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*(3), 267-296.

Ferris, D. (1995). Teaching ESL composition students to become independent self- editors. *TESOL Journal 4*(4), 18-22.

Ferris, D. (1997). The influence of teacher commentary on student revision. *TESOL Quarterly 31*(2), 315-339.

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

Ferris, D. (2002). *Treatment of error in second language student writing*. Michigan, The University of Michigan 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*(1), 49-62.

Ferris, D., Chaney, S., Komura, K., Roberts, B., & McKee, S. (2000). Perspectives, problems, and practices in treating written error. *Colloquium presented at the International TESOL Convention*. Vancouver, B.C.

Ferris, D., & Roberts, B. (2001). Error feedback in L2 writing classes. How explicit does it need to be? *Journal of Second Language Writing 10*(3), 161-184.

Frantzen, D. (1995). The effects of grammar supplementation on written accuracy in an intermediate Spanish content course. *The Modern Language Journal 79*(3), 329-344.

Hajer, M. & Meestringa, T. (2004). *Handboek taalgericht vakonderwijs*. [The handbook of content-based second language instruction.] Bussum, Coutinho.

Hazenberg, S. (1994). *Een keur van woorden. De wenselijke en feitelijke receptieve woordenschat van anderstalige studenten.* [A choice of words. The desired and factual receptive vocabulary of non-native students.] PhD dissertation, Amsterdam, Vrije Universiteit.

Hazenberg, S. & Hulstijn, J.H. (1996). Defining a minimal receptive second-language vocabulary for non-native university students: an empirical investigation. *Applied Linguistics 17*(2), 145-163.

Hedgcock, J. & Lefkowitz, N. (1994). Feedback on feedback: Assessing learner receptivity to teacher response in L2 composing. *Journal of Second Language Writing 3*(2), 141-163.

Hyland, K. & Hyland, F. (2006). Feedback on second language students' writing. *Language Teaching 39*(2), 83-101.

REFERENCES

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*(3), 305-313.

Lalande, J.F. (1982). Reducing composition errors: An experiment. *The Modern Language Journal 66*(2), 140-149.

Leki, I. (1991). The preferences of ESL students for error correction in college-level writing classes. *Foreign Language Annals, 24*(3), 203-218.

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*(1), 43-68.

Robb, T., Ross, S. & Shortreed, I. (1986). Salience of feedback on error and its effect on EFL writing quality. *TESOL Quarterly 20*(1), 83-95.

Sachs, R., & Polio, C.G. (2007). Learners' use of two types of written feedback on a L2 writing revision task. *Studies in Second Language Acquisition 29*(1), 67-100.

Schoonen, R. (2005). Generalizability of writing scores: An application of structural equation modeling. *Language Testing 22*(1), 1-30.

Semke, H. (1984). The effects of the red pen. *Foreign Language Annals 17*(3), 195-202.

Swain, M. (1995). Three functions of output in second language learning. *Principle and practice of applied linguistics: studies in honour of H.G. Widdowson.* G. Cook and B. Seidlhoffer. Oxford, Oxford University Press: 125-144.

Truscott, J. (1996). The case against grammar correction in L2 writing classes. *Language Learning 46*(2), 327-369.

Truscott, J. (1999). The case for "The case against grammar correction in L2 writing classes": A Response to Ferris. *Journal of Second Language Writing 8*(2), 111-122. Truscott, J. (2004). Evidence and conjecture on the effects of correction: A response to Chandler. *Journal of Second Language Writing 13*(4), 337-343.

Truscott, J. (2007). The effect of error correction on learners' ability to write accurately. *Journal of Second Language Writing 16*(4), 255-272.