The effectiveness of comprehensive corrective feedback in second language writing
van Beuningen, C.G.

Citation for published version (APA):
van Beuningen, C. G. (2011). The effectiveness of comprehensive corrective feedback in second language writing Oisterwijk: Boxpress

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
Chapter 3

The effects of direct and indirect corrective feedback on L2 learners’ written accuracy

Small-scale study

3.1 Abstract

Among scholars, there has been continuing disagreement about the benefits of corrective feedback (CF) on second language (L2) learners’ written output. While some researchers advocated the usefulness of CF, Truscott claimed 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. Research outcomes could not yet settle this debate since only the short-term effectiveness of CF could be demonstrated (i.e. the effect of CF on text revisions). Due to methodological shortcomings, results from studies that investigated long-term effects of error correction on learners’ accuracy development 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 study presented in this chapter intended to make a contribution to the ongoing error correction debate. The efficacy of direct and indirect CF was compared to the effects of two control treatments: a treatment that offered learners an extra opportunity to practice their writing skills, and a treatment in which pupils self-corrected their errors without any available feedback. Results showed that CF can be effective in improving learners’ accuracy: while short-term effects were found for both direct and indirect CF, only direct feedback proved to have a significant long-term effect. Neither of the control treatments had a significant effect on pupils’ accuracy development.

3.2 Introduction

As Swain (1995) argued, it is important for teachers to draw on L2 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 allows learners more time and opportunity to compare their 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 claimed that written production and feedback are of special importance for second language acquisition (SLA).

A crucial question is what this feedback should look like. A feedback type commonly used in classrooms is CF: the marking of a student's error by the teacher. In the past decade, 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, argued that CF on L2 learners' writing is not only unnecessary and ineffective, but even counterproductive (Truscott, 1996; 1999; 2004; 2007). He based this claim on two types of arguments. On the one hand, Truscott indicated several theoretical problems associated with error correction. He argued that language teachers – when providing CF – 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 regarded error correction as ineffective on the basis of practical considerations; he doubted whether teachers are capable of providing feedback adequately and consistently, and if so, he still questioned students' ability and willingness to use the received feedback effectively. Based on these objections, Truscott (1996) explained that it should be hardly surprising that earlier studies did not convincingly prove the effectiveness of CF, and concluded that all error correction practices should be abandoned. Truscott (2004) furthermore argued that, until its usefulness has been proven by research findings, CF could 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, made a stand for the use of error correction in writing instruction. In her opinion, Truscott's conclusions were premature. She reasoned that results from prior research had 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 argued that more, well designed research is necessary before any conclusions can be drawn about the (in)effectiveness of error correction in improving students’ accuracy performance in future writing (Ferris, 2002).

Acting upon this call, the study reported on in this chapter investigated the effect of CF on students’ accuracy of both revisions and new pieces of writing, in a tightly controlled set-up with three experimental sessions. We compared the effectiveness of direct CF and indirect CF to two different control treatments. Our study was conducted at Dutch multilingual secondary schools that adopted a language sensitive approach to content instruction. Before elaborating on the present study, we will first take a critical look at earlier research addressing both the questions if and how error correction should be done in L2 (writing) instruction.

### 3.3 Empirical background

#### 3.3.1 Investigating the effectiveness of corrective feedback

While a lot of studies made claims about the (in)effectiveness of error correction, most of them were actually unable to unambiguously interpret their findings, since they lacked a proper control group that received little or no correction. Studies that did include a control group and investigated the short-term efficacy 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 learners who did not receive any CF. In contrast, results from studies investigating the effects of CF on subsequent writing (e.g. Chandler, 2003; Kepner 1991; Polio et al., 1998; Semke, 1984), were 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 CF 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 CF. Therefore, it is feasible that these results could be explained by the differing amounts of writing practice opportunity (i.e. time-on-task) the two treatment groups had to their disposal.
Polio et al. (1998) reported that both students who received CF and students who did not, were able to improve their written accuracy over time. However, students in this study’s CF condition only produced half as many journal entries as the control group did, because of the editing activities they had to perform. Thus, it might well be that any potential advantages of error correction were leveled-out by the beneficial effects of extra writing practice.

Kepner (1991) did not find any significant differences in error-counts between learners who were provided with CF, and students who received content-related comments on an initial piece of writing. However, the flaw in this study – as observed by Chandler (2003) – is that students were not required to do anything with the CF they received. Hence, it remained unclear if learners processed the feedback they were presented with. Since CF could not be expected to be beneficial without being used, the findings from this study does not warrant any conclusions about the (in)efficacy of error correction.

In a two-phase study, Chandler (2003) tried to overcome the methodological shortcomings of the studies described above. In the first phase she addressed “the question of whether error correction can be an effective way to improve accuracy of second language writing” (p. 268). Chandler reported a significant advantage of error correction over a lack of CF. However, as Truscott (2004) pointed out, Chandler’s control group was not an actual control group, and, as a result, she was equally unable to make any judgments on the effectiveness of error correction. The problem was that students in Chandler’s control group did receive CF, but were not asked to revise their writing before the end of the semester (and the end of the data collection). Thus what this study 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 explaining the difference in accuracy gains found between the control group and the experimental group in Chandler’s study.

3.3.2 Direct and indirect corrective feedback
Whereas only a few studies tested an error correction condition against a true control treatment, there are quite some CF studies that examined the relative effectiveness of different feedback types, with the dichotomy between direct and indirect CF receiving the

---

2 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 CF were found. In Chandler’s (2003) 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.
lion’s share of researchers’ attention. While indirect CF 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 corresponding target form.

It has been claimed that L2 learners benefit more from indirect CF 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 CF were inconclusive (e.g. Chandler, 2003; Ferris, 2006; Frantzen, 1995; Lalande, 1982, Rob et al., 1986).

A longitudinal study by Lalande (1982) showed that students who received indirect CF outperformed students in a direct CF group. Frantzen (1995) and Rob et al. (1986), on the other hand, reported that direct and indirect CF were equally effective. A study by Ferris (2006) revealed yet another pattern; whereas indirect correction proved to be most effective in improving L2 students’ accuracy of newly written texts, students who received direct CF made the most accurate revisions. Finally, as opposed to Lalande (1982) and Ferris (2006), Chandler (2003) found that direct CF resulted in the largest accuracy gains, not only in revisions but also in subsequent writing. (See Chapter 2, section 2.5.2, however, for a critical discussion of these studies.)

3.4 Research questions

Earlier research did not provide us with conclusive evidence on the question if and how written CF should be provided. The present study therefore aimed 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:

RQ 1 Does corrective feedback help L2 learners to improve the accuracy of an initial piece of writing during revision?

RQ 2 Does corrective feedback help L2 learners to improve the accuracy of their subsequent writing?
RQ 3  If so: What kind of corrective feedback (i.e. direct CF vs. indirect CF) on L2 learners’ written output is most effective?

3.5 Methodology

3.5.1 Setting and participants
The present study was conducted at two Dutch secondary schools with multilingual student populations. Around 80% of the pupils came from a non-Dutch language background, Arabic and Turkish being the most common L1’s\(^3\). Although most pupils were born in the Netherlands, many of them only started learning Dutch in school (i.e. at age four).

Both schools adopted a language sensitive approach to content teaching. 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 L2 and low language proficiency learners, who might experience problems understanding and acquiring a subject’s 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 of this study consisted of three classes of pupils (N = 66) in their second year of secondary pre-vocational education (or vmbo-t in Dutch). Pupils 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.

3.5.2 Treatments
Four different treatments were included in this study; two experimental treatments: (a) direct CF (hereafter direct) and (b) indirect CF (hereafter indirect), and two control treatments: (c)

\(^3\) Other L1’s within the participant group were (in alphabetic order): Aramaic, Bengali, Berber languages, English, Farsi, Hindi, Kurdish, Spanish, and Urdu.
practicing writing (hereafter *practice*) and (d) revision without feedback (hereafter *self-correction*).

**Experimental treatments: direct and indirect corrective feedback**

Pupils in the direct and indirect treatment groups received comprehensive CF on the texts they produced in the pre-test session (cf. section 3.5.3). All feedback was provided by the same researcher. This is important because earlier research has shown that teachers/researchers may differ in the way they provide CF (Ferris, 2006). Having one person providing all CF instances, thus “ensures greater consistency in treatment and […] enables assessment of the effects of feedback without this potentially confounding variable” (p. 93).

Whereas direct CF took the form of identifying both the error and the target form (cf. example 1), indirect CF only consisted of a code identifying the error and its category (cf. example 2). It was left to the learner to derive the corresponding target form. Learners’ form-related errors were sub-divided 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. A different code was used for each of these nine error types (cf. Appendix C).

**Example 1: direct corrective feedback on form related errors**

Je moet *het* trui *niet* chemisch reinigen. [You should not dry clean the sweater.]

*de* chemisch

**Example 2: indirect corrective feedback on form related errors**

Je moet *het* trui *niet* chemishS reinigen. [You should not dry clean the sweater.]

( __ = wrong word, S = spelling error)

**Control treatments: writing practice and self-correction**

**Writing practice**

Pupils 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,
see section 3.5.4 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. Pupils in the practice group allocated at least as much time to writing as the pupils in the error correction groups. Hence, if we would find accuracy gains for either of the CF treatments but not for the writing practice treatment, we could be sure these gains were brought about by the provision of CF. In this case, Truscott’s (2004) alternative explanation for accuracy differences between treatment groups found in earlier studies, could confidently be ruled out for our own findings.

**Self-Correction**
Participants in the self-correction group were asked to revise their texts without any available feedback, that is to self-correct their errors. We included this treatment to be able to set apart the effects of CF from the effects of revision as such. It may be plausible that learners benefit from having a critical look at their own writing, even without the intervention of a teacher (or researcher in this case).

### 3.5.3 Experimental set-up
The experiment consisted of three sessions (cf. figure 3.1). In the first session (S1), a receptive vocabulary test was administered to establish learners’ overall language proficiency. Furthermore, participants performed the first writing tasks for the two task topics included in the study, that is the metamorphosis of insects and instructions on how to do the laundry (cf. section 3.5.4 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 section 3.5.2 were explained by the researcher. All pupils were also handed a sheet listing the different error types and an example for each error category. They were told that their texts would not only be evaluated with respect to content, but that they also needed to pay attention to form-related issues, such as those on their hand-out.

One week later (S2), pupils 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 (cf. Appendix B). The direct and indirect groups were asked to copy their texts revising all errors the researcher gave feedback on. Pupils in the indirect group were furthermore instructed on the meaning and use of the error
codes in their texts (cf. Appendix C). The practice group was presented with two new writing tasks. The researcher first shortly introduced the topics of these tasks. Pupils 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, learners 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 (S3), 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.

**Figure 3.1** experimental set-up

<table>
<thead>
<tr>
<th>Pre-tests</th>
<th>Treatment</th>
<th>Post-test</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1: week 1</td>
<td>S2: week 2</td>
<td>S3: week 3</td>
</tr>
<tr>
<td>Vocabulary test</td>
<td>Direct CF</td>
<td>Direct feedback</td>
</tr>
<tr>
<td></td>
<td>Indirect CF</td>
<td>Indirect feedback</td>
</tr>
<tr>
<td></td>
<td>Practice</td>
<td>No feedback</td>
</tr>
<tr>
<td></td>
<td>Self-Correction</td>
<td>Revision</td>
</tr>
<tr>
<td></td>
<td></td>
<td>No practice</td>
</tr>
<tr>
<td>Initial writing</td>
<td></td>
<td>No practice</td>
</tr>
<tr>
<td>tasks</td>
<td></td>
<td>Additional practice</td>
</tr>
<tr>
<td></td>
<td></td>
<td>No practice</td>
</tr>
<tr>
<td></td>
<td>Subsequent</td>
<td></td>
</tr>
<tr>
<td></td>
<td>writing tasks</td>
<td></td>
</tr>
</tbody>
</table>

**3.5.4 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\(^4\) (cf. Appendix A). Each series consisted of three tasks, that is an initial task and a subsequent writing task, which were performed by all pupils at S1 and S3 respectively, and an extra task that was performed at S2 by the practice group only\(^5\). 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. Learners were instructed to use at least 15 lines for each

---

\(^4\) (Personal) hygiene matters, such as laundry instructions, are often integrated in Dutch biology methods.

\(^5\) 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.
writing task. Tasks were designed in such a way that the content was unproblematic for all pupils, 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 writing 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 tasks on two topics.

### 3.5.5 Language proficiency pre-test

The instrument we used to obtain an indication of learners’ 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 L2 vocabulary for non-native university students (Hazenberg & Hulstijn, 1996). The original test contains 140 target words, each of which is embedded in a contextually neutral carrier sentence. Pupils 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*’ (cf. Appendix E). 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 learners’ 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; Zareva, Schwanenflugel, & Nikolova, 2005).

### 3.5.6 Dependent measure

Pupils’ accuracy performance was the dependent measure in this study. For all writing tasks, accuracy was calculated as the *number of form-related errors per ten words* (i.e. \((\text{number of form-related errors/total number of words}) \times 10\)).\(^6\) A ratio measure was used to

---

\(^6\) 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.
correct for small individual differences in text length. We used a ten-word ratio instead of the more common hundred-word ratio because the texts pupils produced were relatively short, that is around 120 words. One researcher was responsible for consistently marking the errors in all writing tasks.

Table 3.1 displays the mean accuracy for every treatment group, itemized per session (i.e. S1, S2, and S3), and task topic (i.e. a: metamorphosis, b: laundry instructions). Table 3.2 shows the scores on the receptive vocabulary test per treatment group.

**Table 3.1 Accuracy at S1, S2, and S3**

<table>
<thead>
<tr>
<th>Task Topic</th>
<th>Treatment group</th>
<th>Mean² and standard deviation S1 (pre-test)</th>
<th>Mean² and standard deviation S2 (treatment)</th>
<th>Mean² and standard deviation S3 (post-test)</th>
</tr>
</thead>
<tbody>
<tr>
<td>a: metamorphosis</td>
<td>Direct (N = 15)</td>
<td>1.63 (.76)</td>
<td>0.26 (.32)</td>
<td>1.12 (.74)</td>
</tr>
<tr>
<td></td>
<td>Indirect (N = 15)</td>
<td>1.15 (.53)</td>
<td>0.39 (.44)</td>
<td>1.43 (.78)</td>
</tr>
<tr>
<td></td>
<td>Practice (N = 13)</td>
<td>1.53 (.69)</td>
<td>1.63 (.55)</td>
<td>1.90 (.59)</td>
</tr>
<tr>
<td></td>
<td>Self-Correction (N = 15)</td>
<td>1.21 (.56)</td>
<td>1.12 (.55)</td>
<td>1.32 (.75)</td>
</tr>
<tr>
<td>b: laundry</td>
<td>Direct (N = 18)</td>
<td>1.55 (.79)</td>
<td>0.29 (.30)</td>
<td>1.11 (.60)</td>
</tr>
<tr>
<td>instructions</td>
<td>Indirect (N = 17)</td>
<td>1.24 (.53)</td>
<td>0.60 (.47)</td>
<td>1.29 (.43)</td>
</tr>
<tr>
<td></td>
<td>Practice (N = 14)</td>
<td>1.67 (.45)</td>
<td>1.87 (.54)</td>
<td>1.76 (.91)</td>
</tr>
<tr>
<td></td>
<td>Self-Correction (N = 17)</td>
<td>1.19 (.60)</td>
<td>1.07 (.54)</td>
<td>1.41 (.59)</td>
</tr>
</tbody>
</table>

*Mean number of form-related errors per 10 words

**Table 3.2 Overall language proficiency (score on vocabulary test)**

<table>
<thead>
<tr>
<th>Treatment group</th>
<th>Mean² and standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Direct (N = 18)</td>
<td>65.17 (12.95)</td>
</tr>
<tr>
<td>Indirect (N = 17)</td>
<td>60.35 (13.68)</td>
</tr>
<tr>
<td>Practice (N = 14)</td>
<td>67.07 (14.33)</td>
</tr>
<tr>
<td>Self-Correction (N = 17)</td>
<td>65.94 (13.09)</td>
</tr>
</tbody>
</table>

*Mean number of correct test items out of the 108 total
3.6 Results

3.6.1 Pre-test performance (Session 1)
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 CF, indirect CF, 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 (S1) \(F(3, 54) = 2.10, \ p = .112\), 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) \(F(3, 66) < 1, \ p = .506\).

These results suggest that all treatment groups had a comparable accuracy and L2 proficiency level at the beginning of the data collection. Hence, we can assume that any differences in error counts found later on in the study, are not related to initial differences between treatment groups. Moreover, the task’s topic proved not to have any significant influence on learners’ pre-test accuracy performance.

3.6.2 Accuracy in the revision task (Session 2)
Accuracy at S2 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 S1a (i.e. the number of form-related errors per 10 words committed in the first metamorphosis task), and accuracy S1b (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 S1a and accuracy S1b (i.e. learners’ pre-test accuracy performance) as covariates to account for effects of initial individual accuracy differences.

A significant main effect was found for treatment \(F(3, 51) = 57.57, \ p < .001, \ \eta_p^2 = .77\). 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 Task Topic*Treatment: \(F(1, 51) = 1.40, \ p = \)
While the influence of the language proficiency covariate was non-significant ($F(1, 51) = < 1, p = .905$), accuracy S1a ($F(1, 51) = 5.14, p = .028, \eta^2_p = .09$) and accuracy S1b ($F(1, 51) = 22.21, p < .001, \eta^2_p = .30$) proved to be significant covariates.

Post-hoc pair wise comparisons (using the Bonferroni adjustment) revealed that pupils in the direct group were significantly more accurate (i.e. committed significantly fewer errors) at S2 than pupils in either of the other treatment groups (all $p$-values $\leq .012$). Indirect CF turned out to be the second most effective treatment, being significantly more effective than both practicing ($p < .001$) and self-correction ($p < .001$). Lastly, pupils in the self-correction group outperformed pupils in the practice group ($p = .005$). (Cf. Table 3.3.)

The effectiveness of the different treatments could also be investigated by analyzing accuracy progress between S1 and S2. We did so for the direct, indirect, and self-correction groups. The practice group, however, will be excluded from this analysis, since learners in this group performed two new writing tasks at S2. It would not be valid to directly compare learners’ 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 analyze pupils’ progress between S1 and S3 this way either.

Table 3.1 shows that pupils in the direct, indirect, and self-correction groups gained accuracy between S1 and S2: learners in the direct group reduced the number of form-related errors they committed per 10 words from 1.63 at S1a to 0.26 at S2a, and from 1.55 at S1b to 0.29 at S2b. The number of errors committed by pupils in the indirect group decreased from 1.15 at S1a to 0.39 at S2a, and from 1.24 at S1b to 0.60 at S2b. Error counts for the self-correction group show a reduction from 1.21 at S1a to 1.12 at S2a, and from 1.19 at S1b to 1.07 at S2b.

The progress made by the direct, indirect, and self-correction groups between S1 and S2, was analyzed using a repeated measures ANCOVA with accuracy as the dependent variable, task topic and session 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 session ($F(1, 41) = 16.38, p < .001, \eta^2_p = .29$), and treatment ($F(2, 41) = 3.65, p = .035, \eta^2_p = .15$). There was no significant effect of task topic ($F(1, 41) = < 1, p = .637$). Furthermore, only the interaction between session and treatment ($F(2, 41) = 34.23, p < .001, \eta^2_p = .63$) turned out to be significant. Language proficiency proved to be a significant covariate ($F(1, 41) = 10.41, p = .002, \eta^2_p = .20$). Paired samples t-tests showed that, while the accuracy gains between S1 and S2 were 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 was non-
significant ($p = .092$).

From the above we conclude that pupils 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 learners’ accuracy of an initial text, followed by the self-correction method.
Pupils in these three treatment groups reduced the number of errors they committed
between S1 and S2, but only the accuracy gains of the direct and indirect group showed to
be significant. Providing learners with an extra opportunity to practice their writing, proved
not to have any beneficial effects on their accuracy performance.

Table 3.3 Post-hoc pair wise comparisons S2 based on estimated marginal means

<table>
<thead>
<tr>
<th>(I)Treatment</th>
<th>Direct</th>
<th>Indirect</th>
<th>Practice</th>
<th>Self-Correction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Direct</td>
<td></td>
<td>-0.369a (.113)b</td>
<td>-1.369a (.114)b</td>
<td>-0.954a (.110)b</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$p = .012$</td>
<td>$p = .000$</td>
<td>$p = .000$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$d = 1.22^d$</td>
<td>$d = 4.53^d$</td>
<td>$d = 3.22^d$</td>
</tr>
<tr>
<td>Indirect</td>
<td>-1.000a (.121)b</td>
<td></td>
<td>-0.585a (.108)b</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$p = .000$</td>
<td></td>
<td>$p = .000$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$d = 3.27^d$</td>
<td></td>
<td>$d = 1.95^d$</td>
<td></td>
</tr>
<tr>
<td>Practice</td>
<td></td>
<td></td>
<td>.416a (.116)b</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$p = .005$</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$d = 1.38^d$</td>
<td></td>
</tr>
<tr>
<td>Self-Correction</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*aMean difference (I-J)  
*bStd. Error  
*cSign. (Bonferroni adjusted for multiple comparisons)  
*dEffect size (Cohen’s $d$)

3.6.3 Accuracy in the subsequent writing task (Session 3)

Accuracy at S3 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 S1a and accuracy S1b as covariates.
Results closely resemble those reported above, revealing a significant main effect for
treatment ($F (3, 51) = 5.47$, $p < .005$, $\eta^2_p = .24$), no significant effect of task topic ($F (1, 51) =$
< 1, \( p = .806 \)), nor any significant interaction (all F-values < 1, except Task Topic*Accuracy S1a: \( F(1, 51) = 1.79, p = .187 \)). The language proficiency covariate was again non-significant (\( F(1, 51) = < 1, p = .689 \)), while accuracy S1a (\( F(1, 51) = 6.06, p = .017, \eta^2_p = .11 \)) and accuracy S1b (\( F(1, 51) = 16.23, p < .001, \eta^2_p = .24 \)) proved to be significant covariates.

Post-hoc pair wise comparisons (using the Bonferroni adjustment) revealed that pupils in the direct group committed significantly fewer errors at S3 than pupils in both the practice (\( p = .003 \)) and self-correction group (\( p = .029 \)). Pupils in the direct group also outperformed learners 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 (\( p = 1.00 \)). (Cf. Table 3.4.)

**Table 3.4** Post-hoc pair wise comparisons S3 based on estimated marginal means

<table>
<thead>
<tr>
<th>(J)Treatment</th>
<th>Direct</th>
<th>Indirect</th>
<th>Practice</th>
<th>Self-Correction</th>
</tr>
</thead>
<tbody>
<tr>
<td>(I)Treatment</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Direct</td>
<td>-.408b (153)b</td>
<td>-.579a (154)b</td>
<td>-.437a (148)b</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( p = .061^c )</td>
<td>( p = .003^c )</td>
<td>( p = .029^c )</td>
<td>( p = 1.000^c )</td>
</tr>
<tr>
<td></td>
<td>( d = 1.00^d )</td>
<td>( d = 1.41^d )</td>
<td>( d = 1.09^d )</td>
<td>( d = .07^d )</td>
</tr>
<tr>
<td>Indirect</td>
<td>-.171a (164)b</td>
<td>-.029a (147)b</td>
<td></td>
<td>-.142a (157)b</td>
</tr>
<tr>
<td></td>
<td>( p = 1.000^c )</td>
<td>( p = 1.000^c )</td>
<td>( p = 1.000^c )</td>
<td>( p = 1.000^c )</td>
</tr>
<tr>
<td></td>
<td>( d = .41^d )</td>
<td>( d = .07^d )</td>
<td>( d = .35^d )</td>
<td>( d = .35^d )</td>
</tr>
</tbody>
</table>

\(^a\)Mean difference (I-J)
\(^b\)Std. Error
\(^c\)Sign. (Bonferroni adjusted for multiple comparisons)
\(^d\)Effect size (Cohen’s d)
3.7 Summary and discussion

Results showed that all learners who had the opportunity to revise their writing products (i.e. pupils 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 CF (i.e. direct and indirect) turned out to be significant. Moreover, post-hoc pair wise comparisons revealed that both learners receiving direct CF and pupils whose errors were corrected indirectly, significantly outperformed learners who did not receive any feedback and self-corrected their errors. From this we conclude that it was not revision as such which brought about the improved accuracy of the direct and indirect groups. Thus our first research question, whether or not error correction is an effective means to improve L2 learners’ 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 CF. Whereas short-term efficacy of CF has been defined in terms of accuracy gains during revision, the effects of CF on new pieces of writing has been referred to as a long-term effect (e.g. Ferris, 2002). Adopting these definitions, our findings showed short-term effects for both direct and indirect CF. Direct feedback also proved to have a long-term effect on L2 learners’ accuracy performance. 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 CF, 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 direct CF was superior to the provision of indirect feedback. Firstly, we would like to point out that, even though the difference between the direct and indirect groups was not significant, results did display a trend ($p = .061$), and the p-value was submitted to a strict Bonferroni adjustment. Secondly, as we can see in Table 3.1, pupils in the direct group improved their written accuracy between S1 and S3, while pupils 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 S1a to 1.12 at S3a, and from 1.55 at S1b to 1.11 at S3b. The number of errors committed by pupils receiving indirect feedback, on the other hand, increased from 1.15 at S1a to 1.43 at S3a, and from 1.24 at S1b to 1.29 at S3b. Based on the above, our answer to the second research question – concerning the effect of error correction on newly written texts – is that some, but maybe not all types of error correction result in long-term accuracy gains. Moreover, our careful conclusion in relation to our third research question on the relative effectiveness of
direct and indirect CF, 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 gains.

The fact that direct error correction seemed to be superior to indirect CF when considering its long-term efficacy, contradicts the prediction in the literature that learners benefit more from indirect CF 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 observations also holds for our findings: while pupils who received direct CF could instantly internalize the correct form, pupils 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 effects on learners’ accuracy development, 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 3.1 shows that, whereas direct feedback was effective in improving learners’ accuracy from an initial text to a subsequent writing task, pupils in the indirect group committed more errors at S3 than at S1. The same is true for pupils 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. As was already mentioned in section 3.6.2, learners’ accuracy performance on two different tasks might not be directly comparable because it is uncertain if both tasks are equally difficult. A possible interpretation of the observation that the accuracy performance of learners in the indirect, practice, and self-correction groups decreased between S1 and S3, might therefore be that the S3 tasks were somehow more complex than the initial writing tasks, and that only the beneficial effects of the direct treatment were strong enough to still be observable in spite of this potential difference in task difficulty. An alternative explanation might be found in a reduction of pupils’ motivation as the study went on. Support for this interpretation could be

---

7 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.
found in studies that have shown that learners highly value teacher initiated corrections (e.g. Hedgcock & Lefkowitz, 1994; Leki, 1991). It is conceivable that, in the case of the practice and self-correction treatments, the absence of CF interfered with learners’ motivation. Moreover, pupils who received indirect CF might have been demotivated by the fact that they were uncertain if their own corrections were accurate.

Although our findings suggest that error correction can be an effective means of improving L2 learners’ accuracy in writing, many other questions on issues such as the CF responsiveness of different error types, or the possible avoidance of complex structures due to error correction, remain to be answered by further research.

In his 2007 article, Truscott stated that his case against CF is actually a case against grammar correction. He claimed 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 effectiveness of CF across different error types. Another point of interest is the lexical and/or structural complexity of learners’ 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 claimed that the corrected students in Chandler’s (2003) study, for example, might not have gained accuracy because they benefited from CF, but because they simplified their writing. Truscott argued that it is the immediate goal of CF to make learners aware of the errors they committed, and that this awareness creates a motivation for students to avoid the corrected constructions in future writing (Truscott, 2007). Further research is necessary to test this avoidance hypothesis. Moreover, it has been suggested in the literature (e.g. Ferris, 2004; Hyland and Hyland, 2006; Sheen, 2007) that learners’ levels of L2 proficiency and meta-linguistic awareness might mediate the (relative) efficacy of direct and indirect CF (cf. Chapter 2, section 2.6.3). It would therefore be interesting to investigate the influence of such mediating factors on the efficacy of the different CF forms. These three issues will be addressed in the large-scale follow-up study (N = 268) presented in Chapter 4, which furthermore incorporated a delayed post-test to investigate whether or not the effect of error correction still prevailed four weeks after the moment of CF provision.

Although the questions raised above should be attended to in future research, the findings presented in this chapter clearly showed that the use of CF in content-focused classes has the potential to improve pupils’ written accuracy over time. Moreover, direct

---

8 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.
error correction seemed to be more effective than indirect error correction for this study’s population, since the provision of direct CF did not only lead to more accurate revisions, but also to more accurate performance on new writing tasks.