Corrective feedback (CF) or error correction is a widely used method of targeting linguistic problems in second language (L2) learners’ writing. The role of CF in the process of acquiring an L2 is however an issue of considerable controversy in the field of second language acquisition (SLA). Questions such as if, how, and when CF works, what type of CF is most effective, and which errors are (most) correctable, are yet to be answered by empirical research. The studies presented in this book intend to contribute to the settlement of these issues.

In two quantitative experiments and one qualitative study, Catherine van Beuningen investigates the effects of direct and indirect comprehensive CF on L2 learners’ writing. The studies set out to explore both the value of CF as an editing tool as well as its ability to constitute long-term accuracy improvement. In addition, the studies test whether CF has negative side-effects that harm accuracy development, and look into the influences of potentially mediating factors such as the nature of the targeted error.

In providing robust evidence on the effectiveness of written error correction, the results presented in this thesis suggest that comprehensive CF is a valuable pedagogical tool. The findings also advance the theoretical understanding of the language learning potential of written CF. Therefore, this book is of interest to SLA scholars, writing researchers, and L2 teachers alike.
The effectiveness of comprehensive corrective feedback in second language writing
The effectiveness of comprehensive corrective feedback in second language writing

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Universiteit van Amsterdam
op gezag van de Rector Magnificus
prof. dr. D.C. van den Boom
ten overstaan van een door het college voor promoties ingestelde commissie,
in het openbaar te verdedigen in de Agnietenkapel
op dinsdag 19 april 2011, te 12:00 uur

doort

Catherine Gerardine van Beuningen

geboren te 's-Gravenhage
PROMOTIECOMMISSIE

Promotor: Prof. dr. F. Kuiken
Co-promotor: Dr. N.H. de Jong

Overige Leden: Prof. dr. J. Bitchener
Prof. dr. C.M. de Glopper
Prof. dr. J.H. Hulstijn
Prof. dr. L. Ortega
Prof. dr. G.C.W. Rijlaarsdam
Dr. J.J.M. Schoonen

Faculteit der Geesteswetenschappen
Recently, I was invited to give a short talk about my life as a Ph.D. candidate at a career orientation event for university students. At the end of the evening, one of the attendees came over to me with a huge grin on her face. I asked her why she was so happy. She explained that she had been unsure if pursuing a Ph.D. would be the thing for her, but that this uncertainty had now disappeared because she gathered from my talk that I had very much enjoyed being a Ph.D. student. And indeed, I have! It has been a valuable, inspiring, and fun four years. I want to take this opportunity to thank all the people who, more or less directly, contributed to this memorable experience.

First and foremost, I would like to thank my promotor, Folkert Kuiken, and my co-promotor, Nivja de Jong. I consider myself extremely lucky, having two such highly involved and supportive supervisors. Folkert, it was really good to know that I could always stop by for questions and advice, while at the same time being granted great freedom and confidence. Nivja, thank you for being the perfect role model. I have learned so much from you over the past few years!

Secondly, I am in debt to the members of my scientific committee. I warmly thank John Bitchener, Kees de Glopper, Jan Hulstijn, Lourdes Ortega, Gert Rijlaarsdam, and Rob Schoonen for agreeing to take part in my defence ceremony, and for their invaluable comments on this dissertation’s manuscript as well as on earlier papers. I am grateful for their interest in my work, and for their cordiality when meeting each other (at conferences).

Moreover, I would like to dedicate a word of thanks to Dolly van Eerde, Maaike Hajer, Patrick Snellings, and Ineke Vedder for their support in the earliest stages of the project, as members of my advisory committee.

I am also grateful to all the pupils who participated in the research project, and to their teachers for welcoming me into their classrooms. Without their willingness and cooperation, I would have been unable to write this book. Furthermore, I would like to thank my intern, Laura Kettner, and student assistant, Marjolein Prins, for accompanying me on some of the school visits, and for helping me with transcribing and coding the data.
I am thankful to the Amsterdam Center for Language and Communication (ACLC) for providing me with plenty of opportunities to follow courses, take part in workshops, and attend (international) conferences. I also thank the Dutch Linguistics capacity group, led by Fred Weerman, for giving me the chance to combine the two things I like to do best: research and teaching. Moreover, I have benefited much from being part of the CASLA research group. I really enjoyed the critical discussions, opportunities for presentation try-outs, and having a sounding board for research-related questions and ideas.

Before I started working on my dissertation, people had warned me about the lonely existence of Ph.D. students. Thanks to many (former) colleagues, I never felt that this stereotypical image applied to my situation. I want to thank all my fellow Dutch linguists and ACLC colleagues for sharing lunches, drinks, diners, Sinterklaas-parties, and even singing workshops. Loulou, Suzanne, Marian, Josefien, Marcel, Marjolein, Irene, Maren, and Antje, it was great to have people around who were in the same Ph.D. boat! A special thanks also goes out to Ingeborg, for nice coffees and chats, to Lotte and Nomi, for their warmth and wit, to Wieneke for putting on her dancing shoes, and to Petra for helping out a colleague in despair whenever needed. Sible, my PCH companion, you made working on multilevel modeling fun, and I cannot thank you enough for helping me with the lay-out of this book! Nada and Sonia, it was so good to have you as my room mates in the challenging final stages of the project!

There are a couple of people outside the University of Amsterdam I would like to thank. To begin with, I have met many wonderful people at conferences over the years. Four of them, however, are especially dear to me: Koen, Veerle, and Daphne, thank you for sharing good times! Lieve, I will always remember our trip to California! Secondly, I thank my fellow board members of Anéla for making such a good team!

My dear friends and family I would like to thank for their support and understanding, especially during the final year of the project. Mum, dad, Joke, and Cees, thanks for entertaining Jip when I had to work extra time. Sabine, René, Leontine, Alex, Marieke, Steven, Pauline, Vincent, Kevin, and Annemiek, I promise to become a social creature again from now on! Dearest Floor and Jip, thank you for always being there for me, and for putting things in perspective.

I saved this final paragraph to thank my two dear paranymphs. Marije, having you as a room mate was great in many different ways. It was great because I could always profit from your expertise, it was great because we stimulated each other’s productivity, it was great because we shared the same applied linguistic interests, but most importantly it was great because you have become such a good friend. Petra, my little sister and best friend, it
feels so good to have you by my side on this important day. And of course a big kiss and thank you for making the cover of this book look as nice as it does!
# Table of contents

Acknowledgements

List of figures

List of tables

1  Introduction
  1.1 Linguistic challenges in multilingual classrooms
  1.2 Writing in a second language
  1.3 Corrective feedback (CF) in second language writing
  1.4 Focus of the present research
  1.5 Descriptions and operationalizations of main concepts
    1.5.1 A cognitive perspective on second language development
    1.5.2 Corrective feedback
    1.5.3 Accuracy development
  1.6 Thesis outline

2  Corrective feedback in L2 writing: theoretical perspectives and
   empirical insights
  2.1 Introduction
  2.2 Theoretical foundations of the use of CF in L2 instruction
    2.2.1 L1 and L2 acquisition: different processes, different approaches
    2.2.2 CF as a focus-on-form intervention
    2.2.3 CF as noticing facilitator
    2.2.4 Pushed output and CF

...
# TABLE OF CONTENTS

2.3 Objections against the use of CF in L2 instruction 15
   2.3.1 Implicit and explicit L2 knowledge 16
   2.3.2 Developmental readiness 18
   2.3.3 Potential harmful side-effects of CF 18
2.4 Controversies concerning the use of CF in L2 instruction 19
   2.4.1 Which CF method to use 19
   2.4.2 Which errors to correct 22
2.5 Synthesis and critical review of empirical CF work 23
   2.5.1 Research into the effectiveness of written CF 24
   2.5.2 Research into the differential effectiveness of direct and indirect CF methodologies 26
   2.5.3 Research into the value of CF for different error types 27
   2.5.4 Research into the potential harmful side-effects of CF 28
2.6 Open issues and rationale for the present work 29
   2.6.1 The value of comprehensive CF for L2 acquisition 30
   2.6.2 The differential value of direct and indirect CF 30
   2.6.3 Potential mediating factors in CF efficacy 31
   2.6.4 Measuring written accuracy development 32
2.7 Concluding remarks 34

3 The effects of direct and indirect corrective feedback on L2 learners’ written accuracy: Small-scale study 35
   3.1 Abstract 35
   3.2 Introduction 36
   3.3 Empirical background 37
      3.3.1 Investigating the effectiveness of corrective feedback 37
      3.3.2 Direct and indirect corrective feedback 38
   3.4 Research questions 39
   3.5 Methodology 40
      3.5.1 Setting and participants 40
      3.5.2 Treatments 40
      3.5.3 Experimental set-up 42
      3.5.4 Writing tasks 43
      3.5.5 Language proficiency pre-test 44
      3.5.6 Dependent measure 44
TABLE OF CONTENTS

3.6 Results 46
  3.6.1 Pre-test performance (Session 1) 46
  3.6.2 Accuracy in the revision task (Session 2) 46
  3.6.3 Accuracy in the subsequent writing task (Session 3) 48

3.7 Summary and discussion 50

4 Evidence on the effectiveness of comprehensive error correction in L2 writing: Large-scale study 55
  4.1 Abstract 55
  4.2 Introduction 56
  4.3 Literature review 57
    4.3.1 Research into the effectiveness of written CF 57
    4.3.2 Research into the relative effectiveness of direct and indirect CF 61
    4.3.3 Research into the value of CF for different error types 62
    4.3.4 Research into the potential harmful side-effects of CF 63
  4.4 Rationale of the present study and research questions 64
  4.5 Methodology 66
    4.5.1 Setting and participants 66
    4.5.2 Treatments 67
    4.5.3 Procedure 68
    4.5.4 Writing tasks 71
    4.5.5 Data processing 71
    4.5.6 Linguistic measures 72
    4.5.7 Analysis 73
  4.6 Results 74
    4.6.1 Descriptive statistics 74
    4.6.2 Pre-test performance 79
    4.6.3 Effects of comprehensive CF on written accuracy 79
    4.6.4 Effects of direct and indirect CF on grammatical and non-grammatical errors 83
    4.6.5 Effects of CF on written complexity 84
  4.7 General discussion 85
    4.7.1 The language learning potential of comprehensive CF 85
    4.7.2 Interactions between feedback methodology and error type 86
    4.7.3 The potential detrimental side-effects of CF 87
## TABLE OF CONTENTS

4.7.4  Educational level and CF efficacy  87
4.7.5  Limitations and directions for further research  88
4.7.6  Conclusion  89

5  **An in-depth analysis of the effects of direct and indirect corrective feedback in L2 writing: Qualitative study**  91

5.1  Abstract  91
5.2  Introduction  92
5.3  The present study  95
  5.3.1  Aims  95
  5.3.2  Setting and participants  95
  5.3.3  Treatments and procedure  96
  5.3.4  Writing tasks  97
  5.3.5  Data analysis  97
5.4  Findings  99
  5.4.1  The case of Emre: indirect CF and decreasing error rates  99
  5.4.2  The case of Nathalie: direct CF and decreasing error rates  105
  5.4.3  The case of Mehmet: indirect CF and increasing error rates  109
  5.4.4  The case of Dinesh: direct CF and increasing error rates  114
5.5  Conclusions, discussion, and implications  118

6  **Conclusion**  123

6.1  Introduction  123
6.2  Synthesis of main findings  124
  6.2.1  The value of comprehensive CF as an editing tool  124
  6.2.2  The value of comprehensive CF for L2 acquisition  124
  6.2.3  The relative efficacy of direct and indirect CF  125
  6.2.4  Factors mediating CF efficacy  126
  6.2.5  Potential harmful side-effects of CF  130
  6.2.6  How and when individual learners (fail to) benefit from CF  130
  6.2.7  The value of in-depth accuracy measures  131
6.3  Theoretical implications  131
  6.3.1  CF and accuracy development  132
  6.3.2  CF and attention  133
6.3.3 CF uptake and SLA 134
6.3.4 The required level of CF explicitness 135

6.4 Pedagogical implications 136
6.4.1 The value of written CF for language learning 136
6.4.2 The efficacy and ecological validity of comprehensive CF 137
6.4.3 The importance of revising opportunity 137
6.4.4 The complementary value of different CF methodologies 138
6.4.5 The CF potential within a content-focused context 138

6.5 Limitations and directions for further research 139
6.5.1 Research context 140
6.5.2 Research scope 141
6.5.3 Research methodology 141

6.6 Concluding remarks 142

References 145

Appendices 159
Appendix A: Writing tasks 159
Appendix B: Instructions treatment session 177
Appendix C: Error coding system (indirect CF) 181
Appendix D: Language background questionnaire 183
Appendix E: Vocabulary test 185

English summary 187

Nederlandse samenvatting (summary in Dutch) 195

Curriculum Vitae 203
List of figures

**Figure 1.1 (p. 6)**
A cognitive model of L2 acquisition (adapted from VanPatten & Sanz 1995)

**Figure 3.1 (p. 43)**
Experimental set-up (small-scale study)

**Figure 4.1 (p. 70)**
Experimental set-up (large-scale study)

**Figure 4.2 (p. 77)**
Overall accuracy per treatment and time
## List of tables

<table>
<thead>
<tr>
<th>Table</th>
<th>(p.)</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 1.1</td>
<td>4</td>
<td>Methodologies of providing written corrective feedback (adapted from Ellis, 2009a)</td>
</tr>
<tr>
<td>Table 3.1</td>
<td>45</td>
<td>Accuracy at S1, S2, and S3</td>
</tr>
<tr>
<td>Table 3.2</td>
<td>45</td>
<td>Overall language proficiency (score on vocabulary test)</td>
</tr>
<tr>
<td>Table 3.3</td>
<td>48</td>
<td>Post-hoc pair wise comparisons S2 based on estimated marginal means</td>
</tr>
<tr>
<td>Table 3.4</td>
<td>49</td>
<td>Post-hoc pair wise comparisons S3 based on estimated marginal means</td>
</tr>
<tr>
<td>Table 4.1</td>
<td>72</td>
<td>Average levels of intra-rater and inter-rater agreement for measures of accuracy and structural complexity</td>
</tr>
<tr>
<td>Table 4.2</td>
<td>74</td>
<td>Descriptive statistics: overall language proficiency by educational level and treatment</td>
</tr>
<tr>
<td>Table 4.3</td>
<td>75</td>
<td>Descriptive statistics: overall accuracy by educational level, treatment, and session</td>
</tr>
<tr>
<td>Table 4.4</td>
<td>75</td>
<td>Descriptive statistics: grammatical accuracy by educational level, treatment, and session</td>
</tr>
</tbody>
</table>
Table 4.5 (p. 76)
Descriptive statistics: non-grammatical accuracy by educational level, treatment, and session

Table 4.6 (p. 78)
Descriptive statistics: structural complexity by educational level, treatment, and session

Table 4.7 (p. 78)
Descriptive statistics: lexical diversity by educational level, treatment, and session

Table 4.8 (p. 81)
Summary of significant contrasts between treatment groups

Table 5.1 (p. 98)
Linguistic categories and error types

Table 5.2 (p. 100)
Number and types of errors – Emre

Table 5.3 (p. 107)
Number and types of errors – Nathalie

Table 5.4 (p. 110)
Number and types of errors – Mehmet

Table 5.5 (p. 115)
Number and types of errors – Dinesh
Chapter 1

Introduction

1.1 Linguistic challenges in multilingual classrooms

The connection between language and instruction is an inextricable one. Not only does language serve as the medium of knowledge transfer, but developing pupils' language proficiency is also a pedagogical goal in itself (e.g. Prenger, 2005). The language pupils encounter in school is usually more complex than the language used in informal settings, such as at home (e.g. Cummins, 1991; Schleppegrell, 2001; Snow & Uccelli, 2009). This makes that the linguistic demands learners face in an instructional setting are high, even for monolingual speakers of a given language (e.g. Henrichs, 2010). Yet, for pupils from immigrant backgrounds, for whom the instructional language is not their mother tongue, the linguistic challenges are even greater (e.g. Van Eerde & Hajer, 2008; Van Gelderen et al., 2003).

In the four largest cities of the Netherlands (i.e. Amsterdam, Rotterdam, The Hague, and Utrecht), on average 50% of the primary and secondary school pupils are from non-Dutch language backgrounds (Hartgers, 2007). These children either migrated to the Netherlands early in life, or were born in this country but received little or no Dutch language input before they started attending school at the age of four. Research has shown that these pupils lag behind their native speaker peers in school success (e.g. Driessen, 2009; Gijsberts & Herweijer, 2009; Vallen & Stijnen, 1991). It has been claimed that one of the causes of this difference in overall school performance can be found in pupils' level of Dutch language proficiency (e.g. Gijsberts & Herweijer, 2009; Van Gelderen et al., 2003). Second language (L2) speakers of Dutch have shown to fall behind in a broad range of linguistic domains, such as vocabulary knowledge (e.g. Appel & Vermeer, 1998; Hacquebord, Linthorst, Stellingwerf, & De Zeeuw, 2004; Roeleveld & Béguin, 2009; Verhallen &

---

1 Being the official language, Dutch is the most commonly used language of instruction in the Netherlands.
Schoonen, 1993), reading skills (e.g. Hacquebord et al., 2004; Roeleveld & Béguin, 2009),
grammar knowledge (e.g. Hacquebord, 1989), and writing skills (e.g. Sijtstra, 1997).

The poor school performance of ethnic minority pupils in the Netherlands has
furthermore been attributed to the increased linguistic demands learners are currently faced
with; even mathematics is hardly a matter of numbers and formula’s anymore (e.g. Prenger,
2005; Van den Boer, 2003; Van Eerde & Hajer, 2008). Without the necessary linguistic
repertoire, pupils might fail to understand and acquire the content that is being presented to
them, or fall short in successfully communicating their own knowledge and ideas. A lack of
language proficiency could thus become a stumbling block for content learning (e.g.

From the above it was concluded that – in the multilingual reality of Dutch schools –
there is a strong need for extra attention to language, not only in language-orientated
classes, but also in other classes where the overriding focus is on content (e.g. Prenger,
2005; Van den Boer, 2003; Van Eerde & Hajer, 2008). An instructional paradigm which
addresses this need is the language-sensitive approach to content teaching (e.g. Hajer &
Meestringa, 2004; Van Eerde & Hajer, 2008). Among the spearheads of this approach are
the necessity to provide learners with enough opportunity to engage in productive language
use, and the need for language-related feedback. The present empirical work was
conducted in Dutch multilingual secondary schools adopting this language sensitive
methodology.

1.2 Writing in a second language

One of the options for promoting productive language use in an instructional setting is
requiring pupils to write. Even in one’s native language, writing a good text has been
claimed to be a cognitively demanding task, because one has to simultaneously pay
attention to the text’s content, its structure, its formulation, and its linguistic adequacy (e.g.
Kellog, 1994; Schoonen et al., 2003). In order to create a text that meets the requirements
on all of those aspects, writers have to engage in a constant cycle of planning, formulating,
reading, and revising their text (e.g. Flower & Hayes 1980; 1981; Van den Berg &
Rijlaarsdam, 1996). Producing a text in a language which is not your mother tongue should
be considered even more demanding. Whereas formulating grammatically correct
sentences tends to be an automatized process in one’s first language (L1), this is not the
case in an L2 (e.g. Zimmerman, 2000). As a result, L2 writers are forced to direct more of
their attention to language form, leaving fewer cognitive resources for the other aspects of their writing, such as content or text structure (e.g. Cumming, 2001; McCutchen, 2000; Roca de Larios, Manchón, & Murphy, 2006). One of the tasks L2 teachers thus face, is guiding their students through the difficult process of becoming able writers in the target language. This perspective on L2 writing can be referred to as *learning-to-write*.

While writing in an L2 has been argued to be cognitively challenging, it can at the same time be considered of vital importance to L2 development. Current views on second language acquisition (SLA) implicate that it is not enough to expose learners to abundant L2 input, and that learners also need to be pushed to actively use and produce the target language (e.g. Ellis, 2003; 2005; Skehan, 1995; Swain, 1985; 1995). Learners’ active manipulation of language forms, functions, and concepts is thought to play a crucial role in their language learning process, because output production forces learners to process language more deeply and with more mental effort than is necessary during listening and reading (Van Eerde & Hajer, 2008). Having learners engage in L2 writing can thus be considered a means to learn the language. The focus of this book is on this *writing-to-learn* dimension of L2 writing.

### 1.3 Corrective feedback in second language writing

Both L2 researchers (e.g. Han, 2002; Havranek, 2002; Swain, 1991) and practitioners (e.g. Hajer & Meestringa, 2004) have claimed that, in order for output production to foster L2 acquisition, it should be accompanied by feedback on language form. The most commonly used feedback type targeting linguistic problems is error correction or corrective feedback (CF). CF is a method of providing negative feedback, which could be defined as “any indication to the learner that his or her use of the target language is incorrect” (Lightbown & Spada, 2006, p. 197). There are many different ways of supplying CF on L2 learners’ writing. CF methodologies may vary, for example, with respect to their explicitness, their focus, the person providing the feedback, the feedback medium, etcetera. Table 1.1 supplies an overview of the most common written CF strategies.

Even though most teachers, students, and researchers (e.g. Ferris, 1999) agree that CF on learners’ output has an important place in L2 (writing) instruction, some academics (e.g. Truscott, 1996) have claimed CF to be ineffective and potentially harmful. As a result, the exact role of correction in L2 development has been heavily debated in the field of SLA. Questions such as if, how, and when CF works, what type of CF to use, and which errors to
correct, are yet to be answered by empirical research. The studies presented in this dissertation intended to contribute to the settlement of these issues.

Table 1.1 Methodologies of providing written corrective feedback (adapted from Ellis 2009a)

<table>
<thead>
<tr>
<th>Written CF methodologies</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unfocused or comprehensive CF</td>
<td>All errors in a learner’s text are corrected by the teacher.</td>
</tr>
<tr>
<td>Focused or selective CF</td>
<td>The teacher selects one (or a few) type(s) of error(s) to correct.</td>
</tr>
<tr>
<td></td>
<td>Errors outside the chosen focus domain are left uncorrected.</td>
</tr>
<tr>
<td>Direct CF</td>
<td>The teacher provides both an indication of the errors as well as the</td>
</tr>
<tr>
<td></td>
<td>corresponding target forms.</td>
</tr>
<tr>
<td>Indirect CF</td>
<td>The teacher provides some indication of the errors, but it is left to the</td>
</tr>
<tr>
<td></td>
<td>learner to derive the target forms.</td>
</tr>
<tr>
<td>a. Indication + location</td>
<td>a. Errors are underlined or error codes$^2$ are inserted in the text.</td>
</tr>
<tr>
<td>b. Indication only</td>
<td>b. The number of errors is indicated in the margin, without reference of</td>
</tr>
<tr>
<td></td>
<td>their location.</td>
</tr>
<tr>
<td>Meta-linguistic CF</td>
<td>Learners are supplied with meta-linguistic descriptions of their errors</td>
</tr>
<tr>
<td>Reformulation</td>
<td>The teacher gives a grammatically accurate reformulation of a learner’s</td>
</tr>
<tr>
<td></td>
<td>text, while staying as close to the original as possible.</td>
</tr>
<tr>
<td>Electronic CF</td>
<td>Learners are provided with computer-mediated feedback; in-text hyperlinks</td>
</tr>
<tr>
<td></td>
<td>supply information on the nature of the errors, correct usage, etc.</td>
</tr>
<tr>
<td>Peer feedback</td>
<td>A learner’s text is corrected by a fellow student instead of by the teacher.</td>
</tr>
</tbody>
</table>

1.4 Focus of the present research

The studies reported in Chapters 3, 4, and 5 of this thesis set out to investigate the effectiveness of CF in improving L2 learners’ writing. More specifically, they explored the effect of error correction on learners’ ability to write texts that are linguistically accurate. The

---

$^2$ In his table Ellis (2009a) labelled the use of error codes as meta-linguistic CF. In a different paper (Ellis, 2009b), however, he categorized the coding of errors under indirect error correction. In the present work the latter classification was chosen, since – in my opinion – the fact that learners are expected to use the provided codes to self-correct their errors is the most important property of this CF strategy.
studies were conducted in Dutch multilingual secondary schools within the Randstad area that adopted a language sensitive approach to content instruction. The main issues that are addressed in the empirical chapters of this thesis are summarized below. The empirical and theoretical rationale behind these issues will be discussed in Chapter 2.

1. Is written CF an effective editing tool?
2. Does written CF lead to accuracy development in L2 writing?
3. What type of written CF is most effective?
4. What factors mediate the effectiveness of written CF?
5. Does written CF come with any negative side-effects that harm accuracy development?
6. How and when do individual learners (fail to) benefit from written CF?

1.5 Descriptions and operationalizations of main concepts

This section provides a brief overview of the central concepts in this thesis. Since the main objective of the present empirical work was to explore the role of written CF in the process of acquiring an L2, I will start by introducing the cognitive perspective adopted in this book. Secondly, I will explain how CF and accuracy development were operationalized.

1.5.1 A cognitive perspective on second language development

The present work was framed within a cognitive perspective on second language development. In the cognitive paradigm, learning a second language is described as a process in which learners develop an interlanguage system based on their experience with L2 input (e.g. Han, 2002). Input refers to all the target language material a learner is exposed to, either orally or visually (Gass, 1997). However, not all available language input will result in acquisition. Only the subset of L2 input that has been mentally processed by the learner will contribute to L2 development. This subset is referred to as L2 intake. The interlanguage system that develops through intake then enables L2 learners to produce linguistic output. A learner’s interlanguage is viewed to be independent of both his L1 as well as from the target language system, with its own grammar, lexicon, etcetera (Selinker, 1972). In this perspective, errors in learners’ output should be seen as natural and

---

3 The Randstad is a conglomeration of the four largest cities in the Netherlands; that is Amsterdam, Rotterdam, The Hague, and Utrecht.
indispensable parts of the L2 acquisition process. Figure 1.1 visualizes the language learning process.

**Figure 1.1** A cognitive model of L2 acquisition (adapted from VanPatten & Sanz, 1995)

1.5.2 Corrective feedback
As stated previously, CF could be defined as any indication to a learner that his use of a target language form is incorrect. The correction types under investigation in this thesis are a direct and an indirect form of comprehensive, teacher initiated CF (cf. Table 1.1). In the present work, comprehensive direct CF was operationalized as identification of all existing linguistic errors and provision of the corresponding target forms. Comprehensive indirect CF took the form of identifying every error in a learner’s text by means of providing an error code corresponding to the relevant error category.

1.5.3 Accuracy development
The present work opted to explore the ways in which CF affects L2 learners’ ability to produce target-like output. Therefore, the accuracy of learners’ writing was the main dependent variable in the present research. Accuracy was defined as the target-like use of language forms (i.e. grammar, lexis, and orthography).

The effects of CF on L2 accuracy development were assessed in terms of whether or not learners exhibited differences in their ability to use target-like language in revisions and new pieces of writing, as a function of being subjected to a CF or a non-CF treatment. It needs to be noted that the terms L2 development, acquisition, and learning are used interchangeably throughout this thesis.
### 1.6 Thesis outline

The remainder of this thesis consists of a literature review, three empirical studies\(^4\), a chapter discussing the main findings and implications of the present research, and a short overall summary (both in English and Dutch):

**Chapter 2** draws up the theoretical and empirical framework for the research presented in this thesis. It summarizes the theoretical arguments underpinning the use of CF in L2 classrooms, reviews the controversies surrounding the role of CF in L2 learning, and provides a critical summary of the findings produced by earlier research. The chapter concludes with the open issues that will be addressed in the following empirical chapters.

**Chapter 3** reports the findings of a small-scale study (N = 66) investigating the effect of direct and indirect comprehensive CF on L2 learners’ writing. The study set out to explore the value of CF during revision, as well as its role in L2 accuracy development. The potential influences of a task’s topic and pupils’ language proficiency on the effectiveness of CF were also explored.

**Chapter 4** presents a more extensive investigation (N = 268) into the efficacy of direct and indirect comprehensive error correction. This study again looked into the effects of CF on both learners’ revised and newly written texts. It furthermore included a second post-test to gain more insights into the durability of the effects brought about by error correction. Moreover, the study tested the differential value of CF for grammatical and non-grammatical errors, explored the influence of CF on the structural complexity and lexical diversity of learners’ writing, and examined the interaction between participants’ educational level and CF effectiveness.

In **Chapter 5**, the sequential writing performances of four L2 learners are submitted to an in-depth accuracy analysis. The objective of this qualitative exploration of CF effectiveness was to broaden the understanding on how learners engage with the CF they receive, and

---

\(^4\) There is some overlap within Chapters 3 to 5, because these have all been published or submitted for publication in international journals. The advantage this offers the reader is that each chapter can be read on its own. It furthermore needs to be noted that – since Chapters 3 to 5 have multiple authors – the empirical chapters were written from a ‘we’-perspective, whereas the other chapters were not.
on when correction is (not) beneficial to pupils' accuracy development. The study also set out to provide a more detailed account of the CF responsiveness of different error types.

Chapter 6 presents the overall conclusions distilled from the empirical work reported on in the previous chapters. It also sketches out the theoretical and practical implications of the present findings, and outlines some limitations of the present studies that need to be addressed in future research.

The English and Dutch summaries at the end of this book provide a synopsis of the background and rationale behind the work at hand, the three empirical studies, and their main conclusions and implications.
Chapter 2

Corrective feedback in L2 writing: theoretical perspectives and empirical insights

2.1 Introduction

When reviewing their students’ texts, teachers give feedback on a wide range of issues. They might address the text’s content, the way in which its ideas are presented and organized, the appropriateness of the vocabulary that is used, and so on. The type of feedback that has received most of researchers’ attention, however, is feedback on linguistic errors. Such responses to second language (L2) learners’ non-target-like production have been commonly referred to as instances of corrective feedback (CF) or error correction.

The numerous studies investigating the usefulness of CF (see section 2.5) could be situated at the intersections of two academic disciplines, both with their own theoretical and methodological orientations (e.g. Ellis, 2010; Ferris, 2010; Manchón, 2010; Santos, López-Serrano, & Manchón, 2010; Sheen, 2010a): the field of L2 writing and the domain of second language acquisition (SLA).

Researchers in the field of L2 writing have been mainly interested in the question if and how CF can help students to become more able and self-employed writers (e.g. Chandler, 2003; Ferris, 2006). The predominant focus of studies within this strand of research has been on exploring the role of feedback in the process of developing learners’ editing and revision skills. This perspective has been referred to as the learning-to-write dimension of L2 writing (see e.g. Leki, Cumming, & Silva, 2008 for a review).

Currently, however, the research focus seems to be shifting towards the potential of written CF in aiding learners’ interlanguage development. Arising from a writing-to-learn

---

agenda (e.g. Harklau, 2002; Manchón, 2009; 2010; Ortega, 2009; Santos, et al., 2010), and based within a psycholinguistic and cognitive SLA framework, recent studies have been investigating if receiving and processing written CF can lead to L2 learning (e.g. Bitchener & Knoch, 2010a; Ellis, Sheen, Murakami, & Takashima, 2008; Sheen, 2007). Inspired by oral CF studies, these investigations were tightly controlled in their methodological set-up, and measured the effects of CF by comparing learners’ accuracy performance on pre-tests and (delayed) post-tests. The focus of the present chapter is on this SLA oriented area of CF research.

Even though a number of theoretical SLA insights would predict that written CF can foster L2 development, and despite the fact that the efficacy of oral CF is well documented (see, for example, meta-analyses by Li, 2010, Lyster & Saito, 2010, and Mackey & Goo, 2007), the usefulness of written error correction has been and remains a topic of considerable debate (see particularly Ferris, 1999; 2004; Truscott, 1996; 1999; 2007; Truscott & Hsu, 2008). This chapter intends to review the theoretical arguments underpinning the use of CF in L2 instruction, the objections raised against CF, and the empirical evidence concerning the value of written CF to the SLA process. In the final part, I will describe some of the issues that still need to be investigated, and draw up the rationale behind the empirical work presented in this dissertation.

2.2 Theoretical foundations of the use of CF in L2 instruction

The rationale for expecting that error correction can be beneficial to language learning, rests on various theoretical grounds. Without claiming to provide a comprehensive overview of related theoretical notions and insights, this section will summarize some of the theoretical foundations of the use of CF in L2 classrooms.

2.2.1 L1 and L2 acquisition: different processes, different approaches

Since the early 1970’s a communicative approach to language teaching has been dominating the field of L2 instruction. The communicative paradigm was initiated as a movement away from traditional, structural methods of L2 pedagogy, which focused on teaching isolated linguistic features and grammar rules. Inspired by theories of communicative competence (e.g. Canale & Swain, 1980; Hymes, 1971), communicative approaches aim at developing learners’ ability to use the L2 in realistic, meaningful communication. Key ingredients of this approach are providing learners with abundant
comprehensible input (e.g. Krashen, 1981; 1982; 1985) and creating opportunities to engage in meaningful language use. In doing so, communicative approaches construct an environment that promotes naturalistic acquisitional processes, such as implicit and incidental learning (e.g. Krashen & Terrell, 1983; Long, 1985; Skehan, 1998).

Based on the nativist idea that L1 and L2 acquisition are much alike, some researchers have been advocating a fully naturalistic approach to L2 teaching (e.g. Krashen & Terrell, 1983). Having access to ample comprehensible input was thought to be the necessary and sufficient condition for SLA. Learners were expected to comprehend the available input by inferring its meaning on the basis of linguistic information that is embedded in the communicative context. L2 grammatical competence was believed to emerge automatically, without any need for negative evidence (e.g. Krashen, 1981; 1982; 1985; Schwartz, 1993).

However, it seems fair to state that nowadays the consensus within the field of SLA is that L1 and L2 acquisition are not instances of the same phenomenon; the cognitive processes involved in L1 and L2 acquisition do not fully overlap (Doughty, 2003). Research investigating L2 acquisition in naturalistic settings provided support for this hypothesized difference. Studies in the context of French immersion classrooms in Canada, for instance, found that learners failed to acquire a native-like grammatical competence despite the continuous exposure and practice opportunity the immersion context provided for. Whereas learners were typically shown to develop native-like perceptive skills and fluency, they failed to reach a target-like level of accuracy in production (e.g. Swain, 1991 or see e.g. Lyster, 2007 for a review). Findings as these suggest that, however necessary for SLA, abundant comprehensible input is not a sufficient condition for developing a near-native level of accuracy. Révész (2007) clarified that comprehension and acquisition are not just two sides of the same coin, and that “comprehension may occur in the absence of acquiring linguistic knowledge” (p. 5). It is quite possible for learners to grasp the meaning of a message by relying on contextual information and/or already acquired linguistic awareness. In doing so, they may totally circumvent processing the message’s morphosyntactic encoding.

Ample opportunity for language production does not guarantee learners to be pushed beyond strategic and semantic processing either (Révész, 2007). Even when producing output, L2 learners do not necessarily engage in (full) morphosyntactic processing. They are able to construct a message which is communicatively adequate even when formal accuracy is lacking, or, as Skehan & Foster (2001) put it, “language can work despite poor execution. Its meaning is recoverable even if its form is incorrect” (p. 183). Additionally, in total absence of consideration for L2 accuracy, learners might proceduralize such non-
target-like linguistic solutions to communicative problems, and premature fossilization of errors could be the result (e.g. Skehan & Foster, 2001).

If one starts from the idea that a native-like proficiency on all possible levels, including accuracy, is the ultimate goal of L2 instruction, the conclusion should be that a fully meaning-based approach to L2 instruction does not suffice. Instead, some attention to linguistic form is necessary for learners to be able to progress towards well-formedness in their L2 (e.g. Ellis, 2005; Long, 2000; Long & Robinson, 1998; Norris & Ortega, 2000; Skehan & Foster, 2001). Current communicative methodologies (e.g. task-based approaches, content-based approaches, language sensitive approaches to content teaching) indeed all incorporate some form of grammar instruction.

2.2.2 CF as a focus-on-form intervention

As elucidated in the previous section, it is now widely accepted that effective L2 pedagogy should involve – at least at times – attention to linguistic form. Without it, learners’ accuracy development could be expected to be slower, more difficult, and less successful (Doughty, 2003). A pedagogical intervention that has received considerable attention and which has been advocated in the SLA field (see for example Norris & Ortega, 2000 for a review) is Long’s focus-on-form approach (Long 1991; 1996; 2000; Long & Robinson, 1998). According to Long (2000) focus-on-form involves briefly drawing students’ attention to linguistic elements [...] in context as they arise incidentally in lessons whose overriding focus is on meaning or communication. The temporary shifts in focal attention are triggered by students’ problems with comprehension or production (p. 185).

One of the most crucial characteristics of a focus-on-form intervention is that it is provided within a communicative context. The importance behind this, is explained well by Lyster (2007) in terms of Segalowitz’s (1997; 2000) notion of transfer-appropriate learning. Lyster (2007) rephrased the essence of this concept by stating that “… the kind of cognitive processing that occurs while performing [language] learning tasks should ideally resemble the kind of processing involved during communicative language use” (p. 43). The drawback of decontextualized grammar teaching is that learners will have difficulty transferring the knowledge they have gained from isolated grammar lessons to actual language use in a communicative situation. The focus-on-form approach, on the other hand, caters for learning that is transfer-appropriate.
Whereas Long’s definition implies that focus-on-form episodes are unplanned (i.e. incidental), other researchers (e.g. Doughty & Williams, 1998; Ellis, Basturkmen, & Loewen, 2002) have adopted a broader perspective on what can constitute a focus-on-form episode; in their view focus-on-form can be both planned and unplanned, and reactive as well as preemptive.

One of the pedagogical tools identified as a potent focus-on-form instrument is error correction (e.g. Ellis, 2005). CF is a reactive focus-on-form methodology with the specific value of inducing learners’ attention to form in the context of performing a task, in a personalized, individualized manner. It could be argued that CF on written output is especially promising as a focus-on-form intervention. Whereas oral feedback will inevitably interrupt the communicative flow, learners only have to deal with written feedback after meaning has been communicated (Pollio, Fleck, & Leder, 1998).

2.2.3 CF as noticing facilitator

A second fundamental motivation of the focus-on-form methodology – apart from Segalowitz’s (1997; 2000) notion of transfer-appropriate learning – can be found in Schmidt’s Noticing Hypothesis (Schmidt, 1990; 2001). The concept of noticing combines the two crucial cognitive linguistic notions of attention and awareness (Svalberg, 2007). The Noticing Hypothesis states that subliminal SLA is impossible, and that it is only through conscious attention that input can be converted into intake. Schmidt thus argued that noticing is a necessary condition for language learning.

Another essential role associated with attention, is its ability to make learners aware of “a mismatch or gap between what they can produce and what they need to produce, as well as between what they produce and what target language speakers produce” (Schmidt, 2001, p. 6). This concept has been commonly referred to as noticing the gap (e.g. Schmidt and Frota, 1986). Ellis (1995) used the term cognitive comparison rather than noticing the gap because, in his view, learners also need to notice when their output is the same as the input.

When conscious attention to linguistic form is considered facilitative to or even a prerequisite for interlanguage development, focus-on-form interventions such as CF can be

---

2 Although in psycholinguistic and cognitive SLA accounts the general consensus is that attention is a necessary condition for language learning (e.g. Robinson, 2003; Schmidt, 1990; 1994; 2001; Sharwood Smith, 1993; VanPatten & Cadierno, 1993), there is disagreement on the level of awareness that should be involved (e.g. Robinson, 2003; Schmidt, 2001; Tomlin & Villa, 1994). See, for example, Schmidt, 2001 for an elaborate account of attention related issues in SLA.
expected to support the SLA process (e.g. DeKeyser, 1994; Han, 2002). As Hulstijn and Schmidt (1994) stated, they can be considered cognitive focusing devices for learner attention. In raising learners’ awareness of certain linguistic features, CF enables learners to notice the gaps between their own interlanguage output and the target language input (i.e. the feedback provided). Subsequently, these noticing operations could prompt destabilization and restructuring of learners’ developing interlanguage grammar (e.g. Gass, 1997; Long, 1996).

Adams (2003) furthermore pointed at the advantage of written CF over orally provided feedback. Although both modalities provide learners with the opportunity of noticing mismatches between the target language and their interlanguage system, learners might not (always) be able to make the cognitive comparison in online oral language use. The classic psychological conception of attentional resources is that they are limited (Schmidt, 2001); when presented with an overwhelming number of stimuli at any given moment, the human brain might be unable to attend to them all due to a lack of available processing capacity (Al-Hejin, 2004). Being very demanding on learners’ attentional resources, online language production and orally provided CF might produce such a cognitive overload. In writing, on the other hand, learners have enough time – and therefore cognitive resources – to compare their output with the CF they received, which increases the likelihood of learners noticing gaps in their interlanguage (e.g. Polio et al., 1998; Sheen, 2010a).

2.2.4 Pushed output and CF

Krashen (1989) stated that output is nothing more than a product of already acquired L2 competence. Today however, drawing on Swain’s (1985; 1995) Output Hypothesis, most SLA researchers acknowledge that learner output is a valuable source for acquisition (e.g. Ellis, 2003; 2005; Manchón, 2010; Skehan, 1998). Swain argued that the importance of L2 output lies in the fact that output production pushes learners to process language more deeply (i.e. beyond semantic processing) and with more mental effort than is necessary during listening and reading.

Swain (1985; 1995) specified three specific functions of learner output. Firstly, producing output allows learners to test hypotheses about the target language grammar. Secondly, it may trigger the type of meta-linguistic reflection that is beneficial to interlanguage development.

3 It needs to be noted that not all types of output are expected to serve these functions (e.g. Ellis, 2005). The complexity and length of output resulting from controlled practice exercises, for example, is too limited to be beneficial to interlanguage development.
development. Finally, output has the ability to promote noticing and to push learners’ awareness of the gaps and problems in their interlanguage system.

Swain (1991) added that output by itself does not necessarily serve these functions, and recognized the importance of CF by stating that “if students are given insufficient feedback or no feedback regarding the extent to which their messages have successfully (accurately, appropriately, and coherently) been conveyed, output may not serve these roles” (p. 98). Likewise, other researchers (e.g. Han, 2002) have claimed that learners’ output should be accompanied by CF in order to be beneficial to the language learning process:

[While the focus is on meaning, there is a limit to how much an L2 learner can introspect the sufficiency of his own linguistic resources. Also, even if the learner consciously recognizes at that point what he lacks, there is no guarantee, for various reasons, that he will subsequently be able to tune himself in for a solution in the future input, or even if he is, he may not be able to tell whether what he sees as the potential solution is actually the correct solution. Rather, external feedback [...], I shall argue, may significantly facilitate the fulfillment of the ‘noticing’ function (p. 18).]

Again, it might be argued that the beneficial effects associated with written output and CF, will outweigh those of oral language use and correction. The fact that, in writing, learners are not under such strict time constraints as in online oral language production, makes it more feasible that the beneficial roles of output production in combination with CF are actually realized.

### 2.3 Objections against the use of CF in L2 instruction

While the previous section discussed why CF can be presumed to facilitate SLA, some researchers have stated error correction to be entirely unnecessary and ineffective, or even harmful (e.g. Krashen, 1985; Schwartz, 1993; Truscott, 1996). This claim relies on both practical and theoretical arguments. The practical doubts pertain to teachers’ capacities in providing adequate and consistent feedback, and learners’ ability and willingness to use the feedback effectively (Truscott, 1996). The theoretical case against error correction rests on the claim that CF overlooks important insights from SLA theory. The two main theoretical
issues that informed the objections raised by CF opponents will be discussed here, namely the role of explicit L2 knowledge in the language learning process and Pienemann’s (1989; 1998) Learnability Hypothesis. This section furthermore reviews the hypothesized detrimental side-effects of CF.

2.3.1 Implicit and explicit L2 knowledge
An often addressed issue in the field of instructed SLA is the role of conscious grammar knowledge in becoming a proficient user of the L2. Conscious knowledge about the L2 grammatical system has been widely referred to as explicit or declarative knowledge, and opposed to implicit or procedural knowledge (e.g. Bialystok, 1994; DeKeyser, 1998; Krashen, 1981; see DeKeyser, 2003 for an extensive review). Explicit knowledge denotes a conscious awareness of grammatical rules and the appropriate meta-language for labeling and verbalizing this knowledge (Ellis, 2004). Implicit knowledge, on the other hand, is claimed to be unconscious, non-verbalizable, and rapidly and easily accessible during online language use.

Currently, the assumption is that it is their implicit L2 knowledge that enables learners to communicate spontaneously and fluently. How the type of explicit knowledge resulting from grammar instruction contributes to the SLA process, however, “has been and remains today one of the most controversial issues in language pedagogy” (Ellis, 2005, p. 214). Disagreements concern both the value of explicit knowledge in itself and the connection between explicit and implicit knowledge. This debate is important when exploring the effectiveness of error correction, because CF contestants (e.g. Krashen, 1982; Truscott, 1996) have stated that, if CF yields any L2 knowledge at all, this emerging knowledge could only be explicit in nature.

Opponents to the use of CF in L2 classrooms, such as Krashen (1982), claimed that the benefits of explicit knowledge as such to actual L2 performance are rather limited. In Krashen’s view, learners can only use their explicit L2 knowledge during monitoring (i.e. editing of output after it has been initiated by the acquired system), and not in online language use. In exploring the effect of online planning time on learners’ oral language performance, Yuan and Ellis (2003) found, however, that the available planning time improved the accuracy of learners’ online production. This finding suggests that – if

---

4 This assumption holds for oral L2 use in particular. As Bitchener & Knoch (2010a) stated, “the extent to which learners draw upon explicit and implicit knowledge during the writing process is not known. It is likely, however, that they draw upon both explicit and implicit linguistic knowledge” (p. 4). This point will be further addressed in Chapter 6, section 6.3.1.
provided with enough time – learners are able to access their explicit knowledge online, and therefore the value of conscious L2 knowledge is not restricted to monitor use (Ellis, 2005).

Irrespective of the value of explicit knowledge as such, it may be the case that explicit knowledge aids the development of implicit knowledge. However, those opposing the effectiveness of CF adhere to the position that explicit and implicit knowledge systems are entirely distinct, without an interface connecting them. This view is strongly related to Krashen’s (1981; 1982; 1985) proposed distinction between learning and acquisition. According to Krashen, acquisition of implicit knowledge unfolds unconsciously, whereas learning always involves conscious effort, and can only result in explicit knowledge gains. Since, in his view, internalizing linguistic knowledge takes place in two fundamentally different ways, resulting in two separate knowledge bases, Krashen stated that explicit knowledge could never be converted into implicit knowledge.

From such a non-interface viewpoint (e.g. Krashen, 1985; Schwartz, 1993), the line of reasoning behind the claim that CF does not play a facilitative role in the SLA process is thus as follows: while CF can only result in explicit knowledge, actual language use is totally driven by implicit knowledge. The idea that explicit knowledge will never become implicit, then renders the conclusion that learners' interlanguage system is unsusceptible to CF, or, in Truscott’s (1996) words, that CF will only lead to “a superficial and possibly transient form of knowledge” (p. 345) or ‘pseudolearning’.

Alternative perspectives, however, are possible. Many SLA researchers seem to converge on the position that there is an interface connecting implicit and explicit knowledge bases (e.g. DeKeyser, 1998; Hulstijn, 1995; Hulstijn & Schmidt, 1994; McLaughlin, 1990; Schmidt, 1990; Schmidt & Frota, 1986; Swain, 1985). Drawing on Skill Acquisition Theory (e.g. Anderson, 1982), they propose that the gap between explicit knowledge and language use can be gradually bridged by output practice (DeKeyser, 2003). By practicing language production, L2 learners are enabled to consolidate and automatize their linguistic repertoire (Manchón, 2010). CF is believed to further assist this proceduralization of declarative L2 knowledge (Ellis, 2010).

Other scholars adhere to an intermediate position (e.g. Doughty & Williams, 1998; Ellis, 1997; Long & Robinson, 1998). They see implicit and explicit knowledge as being separated, but argue that explicit knowledge may feed into the intake process by helping learners notice the formal features of the input. From this perspective, CF could be

5 See e.g. DeKeyser (2003) for a comprehensive account of opposing views on the relation between implicit and explicit L2 knowledge.
expected to foster interlanguage development because it facilitates the process of noticing (the gap).

2.3.2 Developmental readiness
Another theoretical objection raised by Truscott (1996) in his case against grammar correction, relates to research into naturalistic SLA and the Natural Order Hypothesis (Krashen, 1981; 1982, 1985). Early studies investigating the acquisition in a naturalistic L2 environment, found that different grammatical features were acquired in a relatively strict, predefined order (e.g. Bailey, Madden, & Krashen, 1974; Dulay & Burt, 1974; Pica, 1983). Such findings suggest that learners will only be able to master linguistic forms in consonance with their own internal learning mechanisms, and not in the sequence imposed by a teacher or L2 syllabus (Corder, 1967).

A similar point was made in Pienemann’s (1989) Teachability or Learnability hypothesis, which suggests that learners will only be able to acquire features for which they show developmental readiness. In Pienemann’s view, features that are beyond a learners’ stage of development will not be teachable because “the acquisition process cannot be steered or modeled just according to the requirements of formal instruction” (1989, p. 57). Truscott (1996) deduced that, for CF to have any effect, teachers should align the CF they provide to a learner’s current level of L2 development. If not, learners will be presented with grammatical structures that they are not yet ready to acquire, and as a result, no intake will take place. It has been pointed out, however, that the current insights and research base concerning developmental sequences is too limited to be useful for teaching practice (e.g. DeKeyser, 1998; Ellis, 1997; Truscott, 1996). This led Truscott (1996) to conclude that provision of aligned CF is not (yet) a feasible objective, and that error correction therefore cannot be expected to be beneficial to SLA.

2.3.3 Potential harmful side-effects of CF
CF opponents have not only stated that error correction is unable to lead to accuracy development, but some even argued that CF can be detrimental to the process of L2 acquisition.

A first reason for claiming that CF should be considered counterproductive, is that, in Truscott’s (1996; 2004) view, the time and energy spent on dealing with corrections (both by teachers and students) could be allocated more efficiently to alternative activities, such as additional writing practice.
Secondly, both Krashen (e.g. 1982) and Truscott (e.g. 1996) suggested that, in making students aware of their errors, CF leads to learner stress and anxiety of committing the same errors in future writing. In their view, this anxiety could make learners avoid the erroneous constructions when writing a new text, resulting in simplified writing. This suggestion that the focus on language form induced by CF might lead to a reduction of the linguistic complexity of learners’ output, is in line with predictions from single-resource, limited capacity models of attention which also expect a trade-off between accuracy and complexity (e.g. Skehan, 1998). Within these models, L2 performance is expected to become more complex when learners are willing and feeling free to experiment with the target language. A focus on accuracy, on the other hand, “is seen to reflect a greater degree of conservatism” in which learners will try “to achieve greater control over more stable [interlanguage] elements” while avoiding extending their L2 repertoire (Skehan & Foster, 2001, p. 191). From a multiple-resource perspective on attention (e.g. Robinson, 2003; 2005), however, linguistic accuracy and complexity are not presumed to be in competition because these two form-related aspects of learner output are thought to be closely related to each other.

2.4 Controversies concerning the use of CF in L2 instruction

The previous sections discussed arguments both in favor and against the use of CF in L2 instruction. Even amongst CF advocates, however, some issues relating to the value of error correction remain divisive. In what follows, I will review two of the most heavily debated issues, which are (1) the differential effectiveness of various CF methodologies, and (2) the amenability of different types of errors to CF.

2.4.1 Which CF method to use

As made explicit in Table 1.1 (p. 4), CF on L2 learners’ writing can take many different forms. It was explained how methodologies of written error correction may vary, for example, with respect to their explicitness, their focus, the person providing the feedback, the feedback medium, and so on. The two dichotomies which have been receiving the lion’s share of researchers’ attention are that between focused and unfocused CF, and the contrast between direct and indirect CF. The following is a synopsis of the different positions that have been advanced in the literature concerning the relative effectiveness of these different CF types.
Focused and unfocused CF

As was already explained, the focused-unfocused dichotomy refers to the comprehensiveness of correction methodologies. The unfocused or comprehensive approach involves correction of all errors in a learner’s text, irrespective of their error category. Focused or selective CF, on the other hand, targets a (number of) specific linguistic feature(s) only (e.g. errors in the use of English articles). Errors outside the focus domain are left uncorrected.

Different predictions have been made with respect to the relative effectiveness of focused and unfocused CF. Ellis et al. (2008), for example, claimed that there are theoretical reasons for expecting the focused approach to be more beneficial to accuracy development than unfocused CF. They stated that learners are more likely to notice and understand corrections when they target a specific (set of) error type(s). The idea that noticing and understanding are essential for acquisition (e.g. Schmidt, 1994; Ellis, 2005), led Ellis et al. (2008) to conclude that focused CF has greater potential to impact accuracy development. Sheen (2007) and Bitchener (2008) furthermore argued that unfocused CF may not be the most effective correction method because L2 learners have a limited processing capacity. They claimed that asking learners to deal with CF which targets a broad range of linguistic features at the same time might produce a cognitive overload, and prohibit feedback processing. As noted earlier, however, this attentional capacity problem might be more prominent in the online processing of oral feedback than in the offline handling of written CF (e.g. Sheen, 2010a).

There are also reasons to question the hypothesized superiority of a focused CF approach. It could be argued that focused CF is rather a form of explicit grammar instruction than a focus-on-form intervention (e.g. Bruton, 2009a). This might make it more difficult for learners to transfer what is learned from the feedback to new writing situations (e.g. Segalowitz, 1997; 2000). Additionally, Ferris (2010) and Storch (2010) noted that, from a practical perspective, only targeting specific error types might not be enough; teachers’ purpose in correcting their pupils' written work is (among other things) improving accuracy in general, not just the use of one grammatical feature. Anderson (2010) furthermore added that “[e]specially for content teachers who still wish to address linguistic concerns alongside regular class content, the use of very focused feedback […] is impractical” (Anderson, 2010, p. 5). Finally, observing that some of their errors have been corrected while others have not, might be rather confusing for students.
**Direct and indirect CF**

The second much discussed contrast is that between *direct* and *indirect* error correction. The main factor distinguishing these two types of CF is the learner’s involvement in the correction process. Whereas direct CF consists of an indication of the error and the corresponding correct linguistic form, indirect CF only indicates that an error has been made. Instead of the teacher providing the target form, it is left to the learner to correct his own errors. Indirect correction methods can take different forms that vary in their explicitness (e.g. underlining of errors, coding of errors).

Various hypotheses considering the relative effectiveness of direct and indirect CF have been put forward, some in favor of direct error correction, others supporting the indirect approach.

On the one hand, it has been suggested that learners will benefit more from indirect CF because they have to engage in a more profound form of language processing when they are self-editing their writing (e.g. Ferris, 1995; Lalande, 1982). In this view, the value of the indirect approach lies in the fact that it “requires pupils to engage in guided learning and problem solving and, as a result, promotes the type of reflection that is more likely to foster long-term acquisition” (Bitchener & Knoch, 2008, p. 415).

Advocates of direct CF (e.g. Chandler, 2003), on the other hand, have claimed that the indirect approach might fail because indirect CF provides learners with insufficient information to resolve complex errors (e.g. syntactic errors). Chandler (2003) furthermore argued that, whereas direct CF enables learners to instantly internalize the correct form as provided by their teacher, learners whose errors are corrected indirectly do not know if their own hypothesized corrections are indeed accurate. This delay in access to the target form might level out the potential advantage of the additional cognitive effort associated with indirect CF. Additionally, Bitchener and Knoch (2010b) suggested that only direct CF offers learners the kind of explicit information that is needed for testing hypotheses about the target language.

It has also been suggested that the relative effectiveness of direct and indirect CF methodologies might be determined by several intervening factors. Some researchers have argued, for example, that indirect CF might be less advantageous to lower proficiency L2 learners, since they lack the level of meta-linguistic awareness that is necessary to self-correct their errors (e.g. Ferris, 2004; Hyland & Hyland, 2006). Secondly, Ferris (1999, 2002) proposed that the effectiveness of different CF methodologies is dependent on the type of error targeted. She claimed that rule-governed errors might make good candidates for self-correction based on indirect CF, but that learners need more explicit information in
the form of direct CF to be able to solve problems that are more idiosyncratic in nature. Ferris (2010) furthermore argued that the goal a teacher tries to achieve by providing CF might influence one’s predictions concerning the differential efficacy of direct and indirect correction. She explained that, when mainly aiming for language learning, direct correction might be considered the most beneficial approach because it provides the kind of efficient and explicit input necessary for acquisition. When opting to help students in developing meta-cognitive, revision, and editing skills, on the other hand, indirect CF methods might prove more useful because they demand a more active form of learner engagement. Finally, Ellis et al. (2008) argued that the success of direct and indirect correction is dependent on the type of knowledge a teacher opts to transfer. They stated that both direct and indirect CF could be expected to foster the reinforcement of already (partially) acquired knowledge, but that the indirect method fails to assist the internalization of new linguistic forms.

2.4.2 Which errors to correct
Apart from theorizing about the most effective CF methodology, researchers have also been concerned with the question which errors to target when providing CF. Various proposals have been advanced in relation to this issue.

Corder (1967), for example, differentiated between errors and mistakes. Errors, in his view, reveal gaps in learners’ interlanguage system, and will therefore be systematic themselves. Unsystematic inaccuracies (i.e. slips of the tongue/pen), on the other hand, arise due to performance failures such as memory limitations. Corder suggested that it is useful to correct learners’ errors but not their mistakes.

Burt (1975; Burt and Kiparsky, 1972) distinguished between global and local errors. He labeled errors that could lead to communication breakdown by interfering with the comprehensibility of the utterance, global errors (e.g. word order errors, lexical errors), whereas minor linguistic violations that do not affect the intended meaning of a message were categorized as local errors (e.g. morphological errors). Hendrickson (1978) recommended teachers to only correct global errors since they impair communication.

Although Krashen (1981; 1982; 1985) denied CF to have any role in L2 acquisition, he stated that CF could have value in enabling learners to monitor their L2 production. However, Krashen noted that this potentially facilitative effect of CF is limited to simple and portable features (e.g. third person –s in English), and that CF should therefore only target this type of errors.
Finally, Ferris (1999; 2002) made a distinction between treatable and untreatable errors. She labeled non-idiomatic or idiosyncratic errors as untreatable (e.g. lexical errors), and categorized errors in patterned and rule-governed features as treatable problems (e.g. article errors). Ferris suggested that CF would be most likely to be successful when directed at treatable inaccuracies.

In fact, problems exist with all of the above proposals, and no clear theoretical basis has been provided for any of them. Ellis (2009) argued, for example, that the dichotomy between errors and mistakes is not as strict as Corder (1967) presented it to be, and stated that “the gravity of an error is to a very considerable extent a matter of personal opinion” (Ellis, 2009b, p. 6). He furthermore noted that there are no theoretical grounds on which teachers or researchers can decide whether an error is simple and portable. The same holds for Ferris’ dichotomy between treatable and untreatable errors.

In my opinion, the only distinction that is readily operationalizable is the contrast between grammatical errors and errors outside the grammatical domain, as proposed by Truscott (e.g. 1996; 2001; 2007). Like Krashen, Truscott predicted CF to have no potential value for the development of grammatical competence. He claimed that CF is unable to affect the rules underlying grammatical errors, and suggested that CF could only be beneficial for errors that “are relatively simple and can be treated as discrete items rather than integral parts of a complex system” (Truscott, 2007, p. 258), such as spelling errors. Interestingly, in applying this discreteness criterion, Truscott’s predictions concerning the amenability of some types of errors are contradicting those of Ferris (1999, 2002). Whereas Truscott (2001) claimed that lexical errors, for example, belong to the most correctible L2 problems because they are relatively discrete, Ferris suggested that it is precisely the idiosyncrasy of lexical errors which makes them less suitable targets for CF.

### 2.5 Synthesis and critical review of empirical CF work

Because the role of written CF in L2 acquisition is still a controversial one, the topic has been and still is attracting a lot of research attention. The present review focuses on the body of empirical studies that opted to add to the writing-to-learn agenda by investigating if CF facilitates learners’ written accuracy development⁶. This section will consecutively

---

⁶ Although the writing-to-learn agenda also involves research on the question how CF contributes to SLA processes such as noticing (e.g. Santos et al., 2010; Storch & Wigglesworth, 2010; Swain & Lapkin, 2002; Qi &
discuss insights concerning (1) the potential of CF in yielding accuracy improvement, (2) the differential effectiveness of direct and indirect correction methodologies, (3) the amenability of different error types to CF, and (4) the possible negative side-effects of error correction.

2.5.1 Research into the effectiveness of written CF

Early research
The earlier work on the effect of CF on L2 learners' written accuracy could be categorized into two strands. While the first set of studies focused on the role of CF during the revision process (e.g. Ashwell, 2000; Fathman & Whalley, 1990; Ferris, 1997; Ferris & Roberts, 2001), the second group of investigations set out to answer the question if correction yields a learning effect (e.g. Chandler, 2003; Kepner, 1991; Polio et al., 1998; Semke, 1984; Sheppard, 1992). The revision studies revealed that CF enables L2 students to improve the accuracy of a particular piece of writing during revision (e.g. Ashwell, 2000; Fathman & Whalley, 1990; Ferris, 1997; Ferris & Roberts, 2001). This finding is valuable from a learning-to-write perspective because it shows that CF has the ability to help learners develop more effective revision and self-editing skills (Ferris, 2010). From an SLA viewpoint, however, such revision studies are less compelling. Polio et al. (1998) already identified development (i.e. the long-term effects of pedagogical interventions such as CF) to be the ultimate concern of SLA research. Truscott and Hsu (2008) were therefore right in claiming that, in only comparing two versions of the same text, the revision studies do not provide evidence of L2 acquisition. Instead, evaluating the potential of CF in yielding a learning effect would necessarily involve “a comparison between two independently written works” (Truscott & Hsu, 2008, p. 293).

Studies that could shed light on the role of error correction in accuracy development are the ones that investigate the effect of CF on new pieces of writing. Earlier studies that did opt to provide insights into the SLA potential of CF, however, rendered inconclusive results (e.g. Chandler, 2003; Kepner, 1991; Polio et al., 1998; Semke, 1984; Sheppard, 1992). Since their conflicting findings could be attributed to methodological issues (such as time-on-task differences or the lack of a control group7), both opponents (e.g. Truscott, Lapkin, 2001), this chapter will only focus on studies investigating the outcome of the learning process (i.e. accuracy development).

7 See Chapter 3, Ferris (2004), or Guenette (2007) for reviews of these design related problems.
THEORETICAL PERSPECTIVES & EMPIRICAL INSIGHTS

1996) and advocates (e.g. Ferris, 1999) of written CF called for more, well designed CF studies.

Recent research
The above mentioned appeal has resulted in a growing body of tightly controlled investigations, exploring the long-term effects of CF on L2 writing, by comparing learners’ accuracy performance on pre-tests and (delayed) post-tests. When considering the type of feedback under investigation, these studies fall into three groups: those evaluating the effectiveness of focused CF, those examining the effects of comprehensive or unfocused correction, and those comparing the efficacy of focused to that of unfocused CF approaches.

Research into the effectiveness of focused CF
Most recent CF research explored the effects of focused CF (e.g. Bitchener, 2008; Bitchener & Knoch, 2008; Bitchener & Knoch, 2009; Bitchener & Knoch, 2010a; Bitchener & Knoch, 2010b; Ellis et al., 2008; Sheen, 2007; Sheen, 2010b). Following the methodology of oral feedback studies (e.g. Lyster, 2004; Ellis, Loewen, & Erlam, 2006), the focused CF approach targets specific, persistently problematic features only (e.g. errors in the use of English articles), leaving errors outside the focus domain uncorrected. These tightly controlled studies all found robust positive effects of focused CF. Moreover, the reported accuracy gains proved to be very durable; Bitchener and Knoch (2010a), for example, showed that students who had received focused CF (only once) continued to outperform students whose errors had not been corrected over a 10-month period.

Research into the effectiveness of unfocused CF
As compared to the growing amount of evidence on the effectiveness of focused CF, proof on the (in)efficacy of comprehensive or unfocused CF is practically non-existent. Only one recent study that I am aware of, aimed at investigating its potential to aid SLA (Truscott & Hsu, 2008).

Truscott and Hsu (2008) contrasted a group receiving unfocused CF with a control group whose errors were not corrected. Truscott and Hsu found that, while comprehensive CF enabled their learners to improve the accuracy of a particular text during revision, it did not lead to accuracy gains in a new text. However, the fact that unfocused CF did not lead

---

8 See Xu (2009), however, for a critical discussion of the findings by Bitchener (2008) and Ellis et al. (2008), and see Bitchener (2009) for a response. See also Storch (2010) for an apprehensive review of recent CF studies.
to learning in this study might have been attributable to a ceiling effect (Bruton, 2009a); the
texts learners wrote during the pre-test held very few errors to begin with, and, as a result,
little room was left for CF invoked improvement in the post-test.

Research into the relative effectiveness of focused and unfocused CF

Little is known about the relative effectiveness of focused and unfocused CF. The only two
studies addressing this issue are Ellis et al. (2008) and Sheen, Wright, and Moldawa
(2009), but results from those two studies contradict each other. Whereas Ellis et al. (2008)
did not find any difference in accuracy gains between their focused and unfocused CF
groups, Sheen et al. (2009) reported an advantage of selective correction over
comprehensive CF. However, both studies suffered from methodological issues as will be
argued in Chapter 4 (see also Xu, 2009 for a critical review of the findings by Ellis et al.,
2008).

2.5.2 Research into the differential efficacy of direct and indirect CF

methodologies

As already explained, different CF methodologies have often been categorized as either
direct or indirect types of correction, and various hypotheses concerning their relative
effectiveness have been put forward. Although a reasonable number of studies opted to
gain insights into the differential effects of direct and indirect CF on learners’ written
accuracy development, empirical findings have not yet been able to confirm any of the
predictions made in the literature. Results from studies that contrasted the effects of direct
and indirect CF (e.g. Chandler, 2003; Ferris, 2006; Frantzen, 1995; Lalande, 1982; Rob,
Ross, & Shortreed, 1986) produced conflicting results. The main cause of this lack of
convergence could be found in design-related and analytical flaws, as will be explained
below.

Lalande (1982), who compared the effect of direct correction to the provision of error
codes (i.e. indirect form of CF) on the accuracy development of 60 learners of German as a
foreign language, reported an advantage of indirect over direct CF. However, the observed
between-group difference in accuracy gains was not statistically significant. Moreover, the
two treatments differed in more respects than just the method of CF provision; the indirect
group was engaged in more form-focused activities than the group receiving direct CF.

Ferris (2006) also reported an advantage of indirect CF over direct correction in
improving the accuracy of 86 ESL students’ writing over time. However, Ferris’ study was
not initially designed to directly compare the two CF methodologies. She set out to explore
the value of indirect CF only, but found that the teachers in her study addressed different
types of errors with different forms of CF. Conclusions on the differential value of direct and
indirect CF for learners’ general accuracy development are therefore unwarranted.

Chandler’s (2003) study investigated the accuracy development of 20 ESL learners
consecutively receiving direct CF and three types of indirect CF (i.e. underlining of errors
with description, underlining only, and description only). She concluded that direct CF was
the most effective approach. However, the reported difference between direct CF and one
type of indirect correction (i.e. underlining only) was not significant. Additionally, the fact that
students received all four feedback types in an only “partial balanced incomplete block
design” (p. 281), makes it difficult – if not impossible – to come to any conclusions on the
differential value of direct and indirect CF methodologies.

Finally, studies by Frantzen (1995) and Robb, Ross, and Shortreed (1986) failed to
find any clear differences in accuracy improvement between groups receiving direct CF and
groups whose errors were corrected indirectly. These studies are therefore equally unable
to provide insights into the relative efficacy of direct and indirect correction.

One tightly controlled study that investigated the effectiveness of two types of direct
correction and one type of indirect CF (Bitchener & Knoch, 2010b), reported an advantage
of direct correction. Whereas in this study direct and indirect CF proved to be equally
effective in improving learners’ accurate use of English articles over a one week period, only
the effect of the two direct CF treatments was still present ten weeks later.

2.5.3 Research into the value of CF for different error types
It has been argued that one could not expect all linguistic errors to be equally amenable to
(the same type of) CF (e.g. Ferris, 1999; Truscott, 1996), because morphological, syntactic,
and lexical errors represent gaps within different domains of linguistic knowledge (e.g.
Schwartz, 1993). Although various hypotheses regarding the CF responsiveness of different
types of errors have been forwarded over the years, the question which errors to correct
remains an empirical one.

A number of studies explored the effects of CF on separate error types, and all
reported differing levels of improvement for different types of errors (e.g. Bitchener, Young,
& Cameron, 2005; Ferris, 2006; Ferris & Roberts, 2001; Frantzen, 1995; Lalande, 1982;
Sheppard, 1992). Ferris (2006), for example, differentiated between five major error
categories (i.e. verb errors, noun errors, article errors, lexical errors, and sentence structure
errors), and found that students receiving CF only realized a significant reduction from pre-
test to post-test in verb errors. Furthermore, Lalande (1982) discerned 12 error types, and
observed that correction only led to a significant decrease in orthographical errors. Bitchener et al. (2005) investigated how CF influenced learners’ accuracy development on three target structures, and found that CF had a greater effect on the accuracy of past simple tense and articles than on the correct usage of prepositions.

These findings show that the type of the error that is targeted, might influence the efficacy of the provided CF. In my view, however, the earlier work has been too heterogeneous (with respect to the types of errors under investigation, CF type, research design and context, etc.) to be able to come to any definitive conclusions on the correctability of different error types.

2.5.4 Research into the potential harmful side-effects of CF

As was explained in section 2.3.3, one of the reasons for CF opponents (e.g. Truscott, 1996) to object against the use of CF in L2 (writing) classes, is that it may lead to simplified writing by triggering learners to avoid situations in which they make errors. These considerations led Truscott (2004; 2007) to propose that accuracy gains found in earlier correction studies (e.g. Chandler, 2003) might well have been attributable to such avoidance and simplified writing instead of to CF.

Few studies have investigated the influence of written CF on linguistic complexity. However, studies that did (e.g. Chandler, 2003; Robb et al., 1986; Sheppard, 1992), could not come to any warranted conclusions because of inadequate methodology and analysis.

Sheppard (1992) – in a study with a fairly small sample size (N = 26) – found that the writing of both his CF group and his control group (who received feedback focusing on idea generation and formulation) became less complex over time. While this decrease in structural complexity (i.e. measured by means of a subordination index) was non-significant for the control group, it reached significance for the students who received CF. Sheppard reported, however, that the post-test performance of both groups on the complexity measure was not significantly different. Furthermore, it remains unclear from the available data if the pre-test writing of the two groups was equally complex. Without knowing if there was any initial between-group difference, no conclusions can be drawn about the effect of CF on written complexity.

Robb et al. (1986) reported a significant positive effect on written complexity of one of their CF treatments (i.e. indirect CF in the form of error codes). However, without a control group who did not receive CF, the reported beneficial effect of indirect CF cannot be taken as evidence against Truscott’s prediction that error correction leads to avoidance.
The same holds for Chandler (2003) who concluded that CF did not affect the complexity of students' writing. Besides the lack of a control group in this study, Chandler's conclusion seems to be problematic for various reasons. She inferred that CF did not trigger simplified writing from the fact that holistic ratings of students' texts did not change significantly over the semester. First of all, one could never be sure if judges took notice of the complexity of students' texts when rating them holistically. Moreover, it seems feasible that raters adjusted their standards in the course of time; what was judged to be a good or bad text at the beginning of a semester might not have been evaluated in the same way months later. Consequently, the fact that holistic ratings did not change cannot be claimed to prove that complexity did not either.

A second argument that led Truscott (1996; 2004) to conclude that CF should be expected to harm L2 learners' accuracy development, is that it diverts time and energy away from more productive aspects of writing instruction. The only study that directly tested this claim by comparing the effects of CF to those of writing practice, is by Sheen et al. (2009). Their results showed that there was a trend ($p = .07$) for the focused CF treatment (targeting English articles) to result in larger accuracy gains than writing practice$^9$. The fact that this difference did not reach significance could well be related to the type of tasks used in this study. Students were asked to rewrite a short narrative after having read it themselves, and having listened to it read out loud by the teacher. As the authors acknowledge, this type of task could be seen as a noticing task which in itself promoted accuracy. Instead of contrasting CF to authentic communicative writing practice, Sheen et al.'s study actually compared two types of form-focused interventions differing in their level of explicitness. It might well be that the observed difference between CF and writing practice would have been greater if the tasks were less inherently focused on linguistic accuracy.

2.6 Open issues and rationale for the present work

In spite of the wealth of empirical studies that investigated the efficacy of written CF and the valuable insights they have provided, there are still many issues that deserve further exploration. This section explains how the three empirical studies reported in this

---

$^9$ The fact that Sheen et al. (2009) did not find unfocused CF to be more beneficial to writing practice, might be related to the unsystematic way in which this kind of feedback was provided (see also Chapter 4). In my opinion, this lack of systematicity makes any conclusion about the efficacy of unfocused CF unwarranted.
The dissertation (cf. Chapters 3, 4, and 5) tried to contribute to some of the unsettled questions surrounding CF.

### 2.6.1 The value of comprehensive CF for L2 acquisition

Although recent studies provided ample proof for the value of focused or selective CF for L2 accuracy development, robust empirical evidence on the long-term effect of unfocused or comprehensive CF is lacking. There are several reasons, however, to consider exploring the learning potential of comprehensive correction important.

To begin with, comprehensive CF seems to be the most authentic feedback methodology. As noted by several researchers (e.g. Anderson, 2010; Ferris, 2010; Hartshorn et al., 2010; Storch, 2010), teachers who provide CF usually opt to improve the overall accuracy of their students’ writing, not just the use of one specific linguistic feature. Moreover, Bruton (2009a) questioned the extent to which the focused CF studies can still be considered to concern genuine L2 writing. In focusing on just one language feature their materials and CF rather seem to constitute written grammar exercises than authentic writing tasks. Xu (2009) addressed a similar point by stating that a clear focus on one grammatical structure may lead learners to consciously monitor the use of that target feature when performing the post-test(s). Finally, both Truscott (2010) and Ferris (2010) argued that the implications that can be drawn from focused CF studies so far are quite narrow, since they all targeted relatively simple linguistic problems (i.e. article errors). It remains to be seen if CF is also able to remedy more complex errors (e.g. word order errors).

While adopting the tightly controlled design of recent investigations into the effectiveness of focused CF (e.g. Bitchener & Knoch, 2010a), the three studies presented in this thesis (cf. Chapters 3 to 5) aimed to explore the ability of unfocused correction to yield a learning effect.

### 2.6.2 The differential value of direct and indirect CF

In spite of the intuitive appeal of the idea that indirect CF is more beneficial to L2 development than direct correction because it engages learners in reflective learning processes (e.g. Ferris, 2006), recent empirical findings seem to point in the opposite direction (Bitchener & Knoch, 2010b). However, the evidence to date is too scarce to lead to any definitive conclusions on the differential efficacy of direct and indirect CF methodologies. This issue deserves further investigation because it is important for teachers to know how their learners’ errors can be treated most effectively. The studies in this book (cf. Chapters 3 to 5) therefore contrasted direct and indirect CF.
2.6.3 Potential mediating factors in CF efficacy

In the previous sections it was already mentioned that the effectiveness of (a specific type of) error correction might be influenced by intervening factors, such as the nature of the targeted error or the goal a teacher pursues by providing CF. The studies presented in this thesis explored the mediating effects of three such factors: (1) error type, (2) a learner’s educational level, and (3) the topic of the writing task a learner receives feedback on.

Error types

Truscott (2001; 2007) stated that if CF has any value for L2 development, this could only be true for simple, non-grammatical errors, but not for errors in grammar. Since research to date did not yet directly or systematically test this hypothesis, the study presented in Chapter 4 investigated the effect of direct and indirect CF on grammatical problems (i.e. syntactic errors) and non-grammatical errors (i.e. spelling errors).

It could be expected, however, that within these two broad domains separate error types still differ in their level of CF responsiveness (e.g. Ferris, 1999; Truscott, 2001). The qualitative study presented in Chapter 5 therefore opted to provide a more detailed account of the amenability of different types of errors to written CF.

Educational level

The study reported in Chapter 4 furthermore set out to explore the potential mediating effect of learners’ educational level on the efficacy of direct and indirect CF. In the first place, this issue is worth investigating from a practical point of view; for teachers it is important to know if a single CF method is equally effective across different levels of education, or if it is the case that learners from different educational levels benefit from different types of correction.

It is theoretically plausible that the (relative) effectiveness of direct and indirect CF is indeed dependent on pupils’ educational level. In the Netherlands, pupils are enrolled in different levels of secondary education based on their performance on a national student assessment (i.e. Cito Eindtoets Basisonderwijs) covering reading comprehension, spelling, technical reading, listening and writing skills, vocabulary, math, social studies, and information processing (Cito, 2010). Since linguistic skills take up an important place in this assessment, it is to be expected that learners in different levels of Dutch secondary education vary in (among other things) their levels of language proficiency and meta-linguistic awareness. This assumption is also supported by the fact that the Dutch national framework of reference for language skills sets different goals for pupils in different levels of
secondary education with respect to listening, reading, speaking, and writing skills and meta-linguistic knowledge (Expertgroep Doorlopende Leerlijnen Taal en Rekenen, 2008).

It has been claimed that factors such as learners’ levels of language proficiency and meta-linguistic awareness influence the degree to which pupils are able to benefit from correction. Sheen (2007), for example, showed that CF was more effective in promoting noticing and understanding when students exposed greater capacity to engage in language analysis. More specifically, the relative effectiveness of direct and indirect CF has also been proposed to be dependent on a learner’s level of (meta-)linguistic competence. It was hypothesized that lower proficiency learners might be unable to correct their own errors based on indirect CF (e.g. Ferris, 2004; Hyland & Hyland, 2006). Considering these predictions in the literature, exploring the influence of pupils’ educational level on CF efficacy might also be interesting from a theoretical perspective.

**Task topic**

Previous research has suggested that it is not just a learner’s level of L2 proficiency which determines the quality of his writing. Other factors, such as a task’s topic, also proved to contribute to a learner’s written performance (e.g. Meuffels & Van den Bergh, 2005; Schoonen, 2005). Specifically, the level of knowledge about, interest in, and familiarity with a certain topic have proven to influence writing performance (e.g. Benton, Sharp, Corkill, Downey, & Khramtsova, 1995; McCutchen, 1986). Another factor that might add to this task-effect is the conceptual complexity of a writing task (see Kuiken & Vedder, 2008 for a discussion about the effect of task complexity on L2 production). It might be that such factors also influence the extent to which a learner benefits from the CF he receives on his performance on a given task, on a given topic. If this would be the case, it would compromise the generalizability of findings yielded by CF studies. Chapter 3 therefore investigated if a task’s topic mediates CF effectiveness.

**2.6.4 Measuring written accuracy development**

When trying to establish the effect of written CF on learners’ accuracy development, the tasks and measures used to quantify the CF invoked improvement are of course of vital importance. Three related issues are worthy of further investigation.

**Realistic writing tasks**

To be able to validly assess the usefulness of CF to L2 learners’ written accuracy development, it is important that the feedback is provided within a realistic writing context.
Long (2007) claimed, for example, that development can only be measured by examining language use during unmonitored production, when learners' focus is on content rather than on language as an object. In the same line, Ellis (2010) argued that free constructed responses constitute “the most valid measure of the effect that CF has on learners’ linguistic competence if this is defined as implicit knowledge (p. 345). The tasks used in earlier CF studies, however, have often been very constrained and artificial, with a clear focus on accuracy. Bruton (2009b) also identified this as one of the problems of previous CF studies (e.g. Bitchener, 2008; Bitchener & Knoch, 2008; Ellis et al., 2008; Sheen, 2007):

The writing tasks are not communicative in the least, since the students are not expected to express a message that has anything personal or new to say to someone else, the audience factor. They do not have clear writing purposes or goals, and there is no reference to the students knowing why they are writing. There […] [are] no criteria, except implicit correctness. (p. 608).

To overcome these task-related drawbacks the tasks used within the present experiments (cf. Chapters 3, 4, and 5) were realistic in set-up and had clear communicative goals.

**In-depth measures**

As did the majority of CF studies to date, the studies presented in Chapters 3 and 4 investigated the effectiveness of written CF by looking at group performances on global accuracy measures (i.e. number of errors per 100 words, number of error-free T-units, etc.) over time. For various reasons, however, it could be considered useful to also take a more detailed look at individual learners’ sequential accuracy performance.

Storch and Wigglesworth (2010), for example, argued that analyzing how learners actually use CF could provide insights on *how* and *when* learners benefit from error correction. Moreover, Bruton (2009a) suggested that detailed, qualitative analysis of learners' writing performance over time might give a more complete and accurate picture of the accuracy gains brought about by CF than the more common global measures. The multiple case-study presented in Chapter 5 therefore adopted the proposed in-depth approach to measuring accuracy development.
CHAPTER 2

Potential harmful side-effects of CF
As explained in section 2.3.3, CF opponents (e.g. Krashen, 1982; Truscott, 1996) proposed that error correction might be expected to harm rather than to foster learners’ interlanguage development. Two arguments have been advanced for this claim: (1) CF was thought to lead to avoidance of more complex language use, and (2) it was argued that it is more effective for teachers to spend their valuable time to extra writing practice than to correction and revision.

Earlier studies have been unable to refute these hypothesized negative side-effects of CF. Hence, the present work set out to investigate the effect of CF on the lexical and structural complexity of learners’ writing (Chapter 4), and also compared the accuracy performance of learners receiving CF to pupils who had an extra opportunity to practice their writing skills (Chapter 3 and 4).

2.7 Concluding remarks
As was explained in the previous section, issues such as if, how, and when comprehensive CF fosters SLA, the relative efficacy of direct and indirect correction, factors mediating CF effectiveness, and the potential harmful side-effects of written error correction, deserve further empirical investigation. The studies presented in the following three chapters (i.e. Chapters 3, 4, and 5) aimed to advance the understanding of these matters.
Chapter 3

The effects of direct and indirect corrective feedback on L2 learners’ written accuracy¹

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
SMALL-SCALE STUDY

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 factor² 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. 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).

<table>
<thead>
<tr>
<th>Example 1: direct corrective feedback on form related errors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Je moet <strong>het</strong> trui niet <strong>chemisch</strong> reinigen. [You should not dry clean the sweater.]</td>
</tr>
<tr>
<td><em>de</em>               <em>chemisch</em></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Example 2: indirect corrective feedback on form related errors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Je moet <strong>het</strong> trui niet <strong>S</strong> reinigen. [You should not dry clean the sweater.]</td>
</tr>
<tr>
<td>( _ = wrong word, S = spelling error)</td>
</tr>
</tbody>
</table>

**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><strong>Vocabulary test</strong></td>
<td><strong>Direct CF</strong></td>
<td><strong>Indirect CF</strong></td>
</tr>
<tr>
<td>Initial writing tasks</td>
<td>Direct feedback</td>
<td>Indirect feedback</td>
</tr>
<tr>
<td></td>
<td>Revision</td>
<td>Revision</td>
</tr>
<tr>
<td></td>
<td>No practice</td>
<td>No practice</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. (number of form-related errors/total number of words) x 10). 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>Mean± and standard deviation</th>
<th>Mean± and standard deviation</th>
<th>Mean± and standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>S1 (pre-test)</td>
<td>S2 (treatment)</td>
<td>S3 (post-test)</td>
</tr>
<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>Mean± 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 ≤ .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 \( (p < .001, \text{using the Bonferroni} \)
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 |
|---------------------------------|--------|--------|--------|
|                                 | Direct | Indirect | Practice | Self-Correction |
| Direct                          |        |         |         |                 |
| Indirect                        |        | -.369\(^a\) (.113)^b | -.1369\(^a\) (.114)^b | -.954\(^a\) (.110)^b |
|                                 |         | $p = .012^c$ | $p = .000^c$ | $p = .000^c$ |
|                                 |         | $d = 1.22^d$ | $d = 4.53^d$ | $d = 3.22^d$ |
| Practice                        |        |         |         |                 |
|                                 |        | -.1000\(^a\) (.121)^b | -.585\(^a\) (.108)^b |               |
|                                 |         | $p = .000^c$ | $p = .000^c$ |           |
|                                 |         | $d = 3.27^d$ | $d = 1.95^d$ |          |
| Self-Correction                 |        |         |         |                 |
|                                 |        |         |        | .416\(^a\) (.116)^b |
|                                 |         |         |         | $p = .005^c$ |
|                                 |         |         |         | $d = 1.38^d$ |

\(^a\)Mean difference (I-J)
\(^b\)Std. Error
\(^c\)Sign. (Bonferroni adjusted for multiple comparisons)
\(^d\)Effect 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 ($F (3, 51) = 5.47, p < .005, \eta_p^2 = .24$), 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 (\( p = 1.000 \)). (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>Direct</td>
<td>-.408a (.153)b ( p = .061^c ) ( d = 1.00^d )</td>
<td>-.579a (.154)b ( p = .003^c ) ( d = 1.41^d )</td>
<td>-.437a (.148)b ( p = .029^c ) ( d = 1.09^d )</td>
<td></td>
</tr>
<tr>
<td>Indirect</td>
<td>( p = 1.000^c ) ( d = .41^d )</td>
<td>( p = 1.000^c ) ( d = .07^d )</td>
<td>( p = 1.000^c ) ( d = .35^d )</td>
<td></td>
</tr>
<tr>
<td>Practice</td>
<td>( p = 1.000^c ) ( d = .142a (.157)b^d )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-Correction</td>
<td>( p = 1.000^c ) ( d = .35^d )</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)
3.7 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 7. 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.
Evidence on the effectiveness of comprehensive error correction in second language writing

Large-scale study

4.1 Abstract

The current study investigated the effect of direct and indirect comprehensive corrective feedback (CF) on second language (L2) learners’ written accuracy (N = 268). The study set out to explore the value of CF as a revising tool as well its ability to constitute long-term accuracy development. In addition we tested Truscott’s (2001; 2007) claim that correction may have value for non-grammatical errors, but not for errors in grammar, and explored the structural complexity and lexical diversity of learners’ writing to see if students are inclined to avoid more complex constructions due to error correction (Truscott, 2007). Results showed that both direct and indirect CF led to improved accuracy, not only during revision but also in new pieces of writing (i.e. texts written during post-test and delayed post-test sessions). Furthermore, a separate analysis of grammatical and non-grammatical error types revealed that whereas only direct CF initiated grammatical accuracy gains, pupils’ non-grammatical accuracy benefited most from indirect CF. Moreover, CF did not result in simplified writing. These findings suggest that comprehensive CF is a useful instrument that teachers can use to help L2 learners improve their written accuracy over time.

---

1 A slightly adapted version of this chapter has been accepted for publication as: Van Beuningen, C.G., De Jong, N. H., & Kuiken, F. (accepted for publication). Evidence on the effectiveness of comprehensive error correction in second language writing. Language Learning.
4.2 Introduction

Error correction or CF is probably the most widely used feedback form in present day L2 classrooms. Its usefulness however, has been a fiercely debated topic ever since Truscott’s 1996 article in which he claimed error correction to be necessarily ineffective and potentially harmful. In the decade that followed, he has repeatedly presented objections with respect to the use of CF in L2 writing classes (Truscott 1996; 1999; 2004; 2007; 2009).

Truscott’s (1996) statement that CF is ineffective relies on both practical and theoretical arguments. His practical doubts pertain to teachers’ capacities in providing adequate and consistent feedback, and learners’ ability and willingness to use the feedback effectively. His argument against the effectiveness of error correction furthermore rests on the claim that CF overlooks important insights from second language acquisition (SLA) theories. First, Truscott denied CF to have any potential merit because of the mismatch between error correction as a simple transfer of information, and the gradual and complex nature of interlanguage development. He furthermore argued that since morphological, syntactic, and lexical knowledge are not all acquired in the same way, no single form of CF can be expected to be effective in all of these domains. In fact, Truscott denied CF to have any potential value for the development of grammatical competence, because – in his view – CF will be unable to affect the rules underlying syntactic and morphological errors. He suggested that CF could only be beneficial for errors that “are relatively simple and can be treated as discrete items” such as spelling errors, but never for grammatical features which are “integral parts of a complex system” (Truscott, 2007, p. 258). Third, and related to the previous argument, Truscott’s view that CF fails to affect the developing interlanguage system, led him to suggest that CF could at best help in developing explicit declarative knowledge, but never the more important implicit procedural knowledge (e.g. DeKeyser, 2003; Ellis, 2004). He deduced that CF could only lead to ‘pseudo-learning’; it might improve learners’ self-editing and revision skills, but could not result in accuracy development. The fourth and final theoretical objection raised by Truscott concerns the so-called ‘readiness problem’ (Pienemann, 1998). He explained that CF can only be expected to have any value when it is aligned to a learner’s current level of L2 development, but that insights into developmental sequences to date are too limited to be useful for teaching practice. From this Truscott concluded that L2 learners are not likely to benefit from CF.

Besides arguing that error correction is ineffective (for the reasons listed above), Truscott (1996; 2004; 2007) stated that correcting students’ writing could even be counterproductive. One of his arguments was that teachers run the risk of making their
students avoid more complex structures when they emphasize learners’ errors by providing CF. Truscott reasoned that it is the immediate goal of error correction 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). Secondly, Truscott (1996; 2004) claimed CF to be a waste of time, and suggested that the energy spent on dealing with corrections – both by teachers and students – could be allocated more efficiently to alternative activities, such as additional writing practice.

Based on the above, Truscott (1996) summoned the abandonment of CF from L2 writing classes, until its usefulness had been proven by empirical research (Truscott, 1999; 2004; 2007). Although Ferris (1999; 2002; 2004) made a stand for the use of written CF and argued that Truscott’s conclusions were premature, she agreed that evidence from well designed studies was necessary before any firm conclusions could be drawn about the (in)effectiveness of error correction. This call has resulted in an ever expanding body of studies exploring the effects of CF on L2 learners’ writing.

Framed within a cognitive perspective on L2 acquisition, the present study reports on a tightly controlled classroom-based quasi-experiment, which included pre-test, treatment, post-test, and delayed post-test sessions, and compared the effects of direct and indirect comprehensive CF on the written accuracy exhibited by 268 secondary school L2 learners during revision as well as in new writing, one and four weeks after the treatment. The study also examined the differential value of CF for grammatical and non-grammatical error types, as well as any putative complexity avoidance effects of the CF treatments.

4.3 Literature review

4.3.1 Research into the effectiveness of written CF

The value of CF for revising and language learning

Earlier empirical work on the effects of CF on L2 learners writing could be categorized into two strands. While the first set of studies focused on the role of CF during the revision process (e.g. Ashwell, 2000; Fathman & Whalley, 1990; Ferris, 1997; Ferris & Roberts, 2001), the second group of investigations set out to answer the question if correction yields a learning effect (e.g. Chandler, 2003; Kepner, 1991; Polio, Fleck, & Leder, 1998; Semke, 1984; Sheppard, 1992).
Students in the revision studies demonstrated that they were able to improve the accuracy of a particular piece of writing, based on the feedback provided (e.g. Ashwell, 2000; Fathman & Whalley, 1990; Ferris, 1997; Ferris & Roberts, 2001). These findings thus showed that CF is a useful editing tool. However, Truscott and Hsu (2008) rightly argued that these results do not constitute evidence of learning. Whereas measuring a learning effect “necessarily involves a comparison between two independently written works” (Truscott & Hsu, 2008, p. 293), the revision studies mentioned above, only compared two versions of the same manuscript. It thus remains unclear if students’ success in using the feedback during revision will subsequently lead to acquisition of the corrected forms.

The more interesting body of research consists of studies that investigated the effect of CF on new pieces of writing. The first studies that did opt to provide insights into the learning potential of CF, however, produced mixed results (e.g. Chandler, 2003; Kepner, 1991; Polio et al., 1998; Semke, 1984; Sheppard, 1992). Whereas Chandler (2003) concluded that CF is an effective means of improving the accuracy of students’ writing over time, Kepner (1991), Semke (1984), Polio et al. (1998), and Sheppard (1992) failed to find an effect of error correction. All of these studies, however, suffered from serious shortcomings in terms of their design, execution, and/or analyses (see Chapter 3 or Guenette, 2007 for a review of these methodological issues). These investigations thus did not allow for any firm conclusions on the role of CF in L2 accuracy development, and led both opponents (e.g. Truscott, 1996) and advocates (e.g. Ferris, 1999) of written error correction to call for more, well designed CF studies.

The language learning potential of focused and unfocused CF

The above mentioned appeal has resulted in a growing body of tightly controlled investigations, exploring the long-term effects of CF on L2 writing, by comparing learners’ accuracy performance on pre-tests and (delayed) post-tests (see Chapter 3; see also Bitchener, 2008; Bitchener & Knoch, 2008; Bitchener & Knoch, 2009; Bitchener & Knoch, 2010a; Bitchener & Knoch, 2010b; Ellis, Sheen, Murakami, & Takashima, 2008; Sheen, 2007; Sheen, 2010b; Truscott & Hsu, 2008). When considering the type of feedback under investigation, these studies fall into three groups: those evaluating the effectiveness of focused CF, those examining the effects of unfocused CF, and those comparing the efficacy of focused and unfocused CF approaches. The focused-unfocused dichotomy refers to the comprehensiveness of CF methodologies. Whereas the unfocused or comprehensive approach involves correction of all errors in students’ texts, independent of
their error category, focused or selective CF targets a (number of) specific linguistic feature(s) only. Errors outside the focus domain are left uncorrected.

Research into the effectiveness of focused CF
As opposed to the earlier CF work discussed in the previous section, and following the methodology of oral feedback research (e.g. Lyster, 2004; Ellis, Loewen, & Erlam, 2006), the majority of recent written CF studies explored the effects of focused correction on learners’ accuracy development (e.g. Bitchener, 2008; Bitchener & Knoch, 2008; Bitchener & Knoch, 2009; Bitchener & Knoch, 2010a; Bitchener & Knoch, 2010b; Ellis et al., 2008; Sheen, 2007; Sheen, 2010b). In these studies, CF targets one persistently problematic error type only (e.g. errors in the use of English articles). The rationale behind this approach is that learners might be more likely to notice and understand corrections when just one feature is targeted (Ellis et al., 2008). Moreover, it was hypothesized that focused CF should be expected to be more effective than unfocused CF, because learners have a limited processing capacity (Bitchener, 2008; Sheen, 2007). The proposition is that, asking students to deal with CF which targets a broad range of linguistic features at the same time, may produce a cognitive overload, and prohibit feedback processing. Studies investigating the effectiveness of a focused approach to error correction, all reported robust and durable positive effects of CF on learners’ accuracy development (see Xu, 2009, however, for a critical discussion of the findings by Bitchener, 2008, and see Bitchener, 2009 for a response).

Research into the effectiveness of unfocused CF
As compared to the growing amount of evidence on the efficacy of focused CF, proof on the language learning potential of comprehensive or unfocused CF (i.e. correction of every error in students’ writing) is scarce. Notwithstanding the significant contribution of the focused CF work to the error correction debate, we still consider it important to explore the effect of unfocused CF on new pieces of writing. To begin with, the unfocused approach most closely resembles the correction method used in actual teaching practice; a teacher’s purpose in correcting his/her pupils’ written work is (among other things) to improve accuracy in general, not just the use of one grammatical feature (Ferris, 2010; Storch, 2010). This is why Hartshorn et al. (2010) called for research on a more authentic CF methodology, which focuses “on the accurate production of all aspects of writing, simultaneously” (p. 89). Moreover, Bruton (2009a) questioned the extent to which focused CF studies could still be considered to concern genuine L2 writing. He reasoned that, in just
focusing on one linguistic feature, their materials and CF rather seem to constitute written grammar exercises instead of authentic writing tasks. Xu (2009) addressed a similar point by stating that such a clear focus on one grammatical structure, may lead students to consciously monitor the use of that target feature when performing the post-test(s). Finally, the implications that can be drawn from focused CF studies so far are rather limited, because the targeted linguistic feature (i.e. articles) was selected for maximal simplicity (Ferris, 2010; Truscott, 2010).

Only two recent studies that we are aware of, aimed at investigating if unfocused or comprehensive CF yields a learning effect (cf. Chapter 3; Truscott & Hsu, 2008). Truscott and Hsu (2008) examined the writing performance of 47 ESL learners, half of whom received comprehensive corrections while the other half functioned as a control group. Truscott and Hsu found that, while comprehensive CF enabled their learners to improve the accuracy of a particular text during revision, it did not lead to accuracy gains in a new text. However, the fact that unfocused CF did not lead to learning in this study might have been attributable to a ceiling effect (Bruton, 2009a); the texts learners wrote during the pre-test held very few errors to begin with, and, as a result, little room was left for CF invoked improvement in the post-test. The second study into the long-term effectiveness of comprehensive CF was conducted as a pilot for the present study (cf. Chapter 3). It explored the effects of two types of comprehensive CF and two control treatments on the writing of 66 L2 learners of Dutch, and indicated that comprehensive error correction can be beneficial to the SLA process. Findings showed that comprehensive error correction not only led to improved accuracy in the revised version of a particular piece of writing, but that it also yielded a learning effect; learners who received unfocused CF made significantly fewer errors in newly produced texts than pupils whose errors had not been corrected.

Research into the relative effectiveness of focused and unfocused CF

Although hypotheses about the relative efficacy of focused and unfocused CF have been forwarded in the literature, there is little empirical evidence on the differential value of the two feedback methodologies. The only two studies addressing this issue are one by Ellis et al. (2008), and one by Sheen, Wright, and Moldawa (2009). Ellis et al. (2008) did not find any difference in accuracy gains between their focused and unfocused CF groups. However, this study has several methodological weaknesses (see Xu, 2008 for a discussion). One of the problems the authors themselves mentioned, is that students in the focused group received more feedback on the target feature (i.e. articles) than students in the unfocused group. Sheen et al. (2009) found the focused approach to be more beneficial
than provision of comprehensive feedback. However, as acknowledged by the authors themselves, the CF received by the unfocused group was rather unsystematic in nature; while some errors were corrected, others were ignored. It is conceivable that this unsystematic way of correcting has negatively influenced the effect of unfocused CF in this study. Students might have been confused noticing that some of their errors were disregarded.

4.3.2 Research into the relative effectiveness of direct and indirect CF

CF researchers have not only shown interest in the question if correction should be comprehensive or selective in nature. Many studies have also been exploring the relative effectiveness of different CF types. Most of the studies that compared different forms of CF categorized their CF methodologies as either direct or indirect. Whereas direct CF consists of an indication of the error and the corresponding correct L2 form, indirect CF only indicates that an error has been made. Instead of the teacher providing the target form, it is left up to the learner to correct his/her own errors. Indirect correction methods can take different forms that vary in their explicitness (e.g. underlining of errors, coding of errors).

Various hypotheses concerning the relative effectiveness of direct and indirect CF have been put forward, some in favor of direct error correction, others supporting the indirect approach. On the one hand it has been suggested that learners will benefit more from indirect CF because it engages students in a more profound form of language processing as they are self-editing their writing (e.g. Ferris, 1995; Lalande, 1982). In this view, the indirect approach “requires pupils to engage in guided learning and problem solving and, as a result, promotes the type of reflection that is more likely to foster long-term acquisition” (Bitchener & Knoch, 2008, p. 415). Advocates of direct CF (e.g. Chandler, 2003), on the other hand, claimed that the indirect approach might fail because indirect CF provides learners with insufficient information to resolve complex linguistic errors (e.g. syntactic errors). Chandler (2003) furthermore argued that, whereas direct CF enables learners to instantly internalize the correct form, learners whose errors are corrected

---

2 Sheen et al. (2009) justified the unsystematic nature of the unfocused CF treatment in their study by arguing that, in reality, teachers also tend to provide comprehensive feedback in an unsystematic and confusing manner. Even if this is the case, however, it would not validate doing the same for research purposes. What Sheen et al. (2009) have been comparing is not focused and unfocused CF perse, but focused CF and an inferior form of unfocused CF. The fact that the focused approach proved to be more beneficial than comprehensive correction in this study, does not warrant the conclusion that unfocused CF is necessarily less effective. It only shows that high quality focused feedback is more useful than badly provided unfocused CF.
indirectly do not know if their own hypothesized corrections are indeed accurate. This delay in access to the target form might level out the potential advantage of the additional cognitive effort associated with indirect CF. Moreover, it has been suggested that learners need a certain level of (meta-)linguistic competence to be able to self-correct their errors using indirect CF (e.g. Ferris, 2004; Hyland and Hyland, 2006; Sheen, 2007).

Neither of the above hypotheses could yet be confirmed by empirical findings, since clear evidence on the differential effects of direct and indirect CF on accuracy development is lacking. Research on the issue has produced conflicting results (cf. Chapter 2, section 2.5.2). Studies by Frantzen (1995) and Robb, Ross, and Shortreed (1986) failed to find any clear efficacy differences between direct and indirect types of correction. Ferris (2006) and Lalande (1982), on the other hand, reported an advantage of indirect over direct CF. However, Ferris’ study was not initially designed to directly compare the two CF methodologies, and in Lalande’s study, the observed between-group difference in accuracy gains was not statistically significant. Moreover, Lalande’s direct and indirect treatments differed in more respects than just the method of CF provision; the indirect group was engaged in more form-focused activities than the group receiving direct CF. Three other studies concluded that direct correction was the most effective approach (cf. Chapter 3; Bitchener & Knoch, 2010b; Chandler, 2003). In Chandler’s (2003) study, however, the observed difference between direct CF and indirect correction was not significant. Additionally, the fact that students in this study received both direct and indirect CF in an only partial balanced design, makes it difficult – if not impossible – to come to any conclusions on the relative value of different CF types. In the study piloting the present research methodology (cf. Chapter 3), the difference between direct and indirect CF treatments did not reach significance either, at a p-value of .06. However, learners receiving direct CF significantly outperformed pupils in the control groups when writing a new text, whereas this study’s indirect CF group did not. Finally, Bitchener and Knoch’s (2010b) study did report a statistically significant difference between direct and indirect CF, in favor of the direct approach.

4.3.3 Research into the value of CF for different error types
In his 2007 article, Truscott explained that his case against CF (Truscott, 1996, 1999) was actually a case against grammar correction. He claimed that syntactic errors in particular might not be amenable to correction, because they are integral parts of a complex system which – in Truscott’s view – is impermeable to CF. He furthermore suggested that morphological features are evenly unlikely to benefit from CF because their acquisition not
only depends on the understanding of form, but also of meaning and use in relation to other words and portions of the language system. Truscott (2001; 2007) concluded that if CF has any value for L2 development, this could only be true for “errors that involve simple problems in relatively discrete items” (Truscott, 2001, p. 94) – such as spelling errors – and not for errors in grammar.

A number of studies (e.g. Bitchener et al., 2005; Ferris, 2006; Ferris & Roberts, 2001; Frantzen, 1995; Lalande, 1982; Sheppard, 1992) explored the effects of CF on separate error types, and all reported differing levels of improvement for different types of errors (cf. Chapter 2, section 2.5.3). None of these studies, however, could test Truscott’s claim that grammar correction is ineffective, because they did not explicitly and systematically investigate whether CF is more beneficial for non-grammatical error types than for errors in grammar.

4.3.4 Research into the potential harmful side-effects of CF

One of Truscott’s (2004; 2007) reasons to object against the use of CF in L2 writing classes, is that it may lead to simplified writing by triggering learners to avoid situations in which they make errors. These considerations led Truscott to propose that accuracy gains found in earlier correction studies might well be attributable to such avoidance and simplified writing instead of to CF. Truscott’s suggestions are in line with limited capacity models of attention which also predict a trade-off between accuracy and complexity (e.g. Skehan, 1998). Within these models L2 performance is expected to become more complex when learners are willing and feeling free to experiment with the target language. A focus on accuracy, on the other hand, “is seen to reflect a greater degree of conservatism” in which learners will try “to achieve greater control over more stable [interlanguage] elements” while avoiding extending their L2 repertoire (Skehan & Foster, 2001, p. 191).

Few studies have investigated the influence of written CF on linguistic complexity, and in our opinion, studies that did (Chandler, 2003; Robb et al., 1986; Sheppard, 1992), could not come to any warranted conclusions. Sheppard (1992), for example, reported a negative effect of CF on the structural complexity of learners’ writing, but in fact, his finding was non-significant. Robb et al. (1986) found CF to have a significant positive effect on written complexity, but this study did not include a control group who did not receive CF. The same holds for Chandler’s (2003) study, who did not find any effect of CF on the complexity of students’ writing. An additional problem with the latter study is that Chandler based her conclusion on holistic ratings. In our view, however, the fact that holistic ratings did not change does not necessarily prove that complexity did not change either. (See
Chapter 2, section 2.5.4 for a more elaborate discussion of findings by Chandler, 2003; Robb et al., 1986; Sheppard, 1992.)

A second argument that led Truscott (1996; 2004) to conclude that CF should be expected to harm L2 learners’ accuracy development, is that it diverts time and energy away from more productive aspects of writing instruction. The only study that directly tested this claim by comparing the effects of CF to those of writing practice, is one by Sheen et al. (2009). Their results opposed Truscott’s claim by showing that learners did not benefit more from writing practice than from CF.

4.4 Rationale of the present study and research questions

Based on the empirical evidence presented in the previous sections, the error correction debate cannot be fully settled. The present study – of which the set-up, tasks, and procedure were piloted (cf. Chapter 3) – intended to add to the existing body of research by trying to tackle some of the unsettled issues, and answering eight research questions, which will shortly be introduced.

A number of early CF studies showed that comprehensive or unfocused CF enables learners to improve a particular piece of writing through revision. Moreover, there is now growing evidence on the contribution of focused CF to the long-term acquisition of certain linguistic features. More empirical evidence is needed, however, before any definitive conclusions can be drawn about the value of unfocused or comprehensive CF for accuracy development or learning. Hence, the present study aimed to explore the value of comprehensive correction during revision, as well as its ability to yield a learning effect, while adopting the tightly controlled methodology of recent focused CF studies (e.g. Bitchener & Knoch, 2010a):

RQ 1 Is comprehensive written CF useful as an editing tool, in that it enables learners to improve the accuracy of an initial text during revision?

RQ 2 Does comprehensive written CF yield a learning effect, in that it leads to improved accuracy in new texts written one week and four weeks after CF has been provided?

Because it seems plausible that learners benefit from taking a critical look at their own text and revising it, even without teacher intervention, the present study furthermore opted to test if comprehensive CF has an added value above self-correction:
LARGE-SCALE STUDY

RQ 3 Is comprehensive CF more beneficial to learners’ accuracy development than having the opportunity to correct their own writing?

A second issue on which CF research to date was unable to come to any clear conclusions, is the relative efficacy of direct and indirect CF. The present study therefore set out to investigate the differential value of these two CF methodologies:

RQ 4 Which feedback methodology is most effective: direct or indirect CF?

Moreover, since no earlier research directly addressed Truscott’s (2001; 2007) claim that CF might have value for non-grammatical errors, but not for errors in grammar, we strived to test this hypothesis by distinguishing between grammatical and non-grammatical error types in our analyses:

RQ 5 Are grammatical errors less correctable than other types of errors (i.e. non-grammatical errors)?

In addition, very few studies have explored the potential harmful effects of CF. The present study therefore compared the effect of comprehensive correction to the effect of mere writing practice to determine if CF is a waste of time (Truscott, 1996; 2004). Moreover, we examined the lexical and structural complexity of learners’ output to see if pupils are inclined to avoid more complex structures due to error correction (Truscott, 2004; 2007):

RQ 6 Is comprehensive CF more beneficial to learners’ accuracy development than writing practice?
RQ 7 Does error correction lead to avoidance of lexically and structurally (more) complex utterances?

Finally, we explored the potential influence of learners’ educational level on the degree to which they are able to benefit from direct and indirect CF. As was explained in Chapter 2 (section 2.6.3), exploration of this issue is primarily interesting from a practical perspective, but may also lead to theoretical implications. In the first place, it would be valuable for teachers to know if learners from different educational levels are equally receptive to (direct and indirect) CF. Moreover, we presumed learners’ educational level to be indicative of their level of meta-linguistic awareness (cf. Chapter 2 for the rationale...
behind this assumption). The hypothesis that learners with lower levels of (meta-)linguistic competence might be less able to correct their own errors based on indirect CF (e.g. Ferris, 2004; Hyland & Hyland, 2006; Sheen, 2007), then led us to expect that pupils with a higher educational level would profit more from indirect CF than pupils from a lower level of education:

RQ 8  What (if any) is the influence of pupils' educational level on CF efficacy?

4.5 Methodology

4.5.1 Setting and participants
Four Dutch secondary schools with multilingual student populations participated in the study. Over 80% of those schools' pupils came from non-Dutch language backgrounds; although most pupils were born in the Netherlands, many of them only started learning Dutch in school (i.e. at age four). Our sample was very heterogeneous with respect to language background (28 L1’s), Moroccan Arabic (31%), Turkish (16%), and Surinamese languages (16%: 8% Sranan Tongo, 8% Sarnami Hindustani) being the most common L1’s.

All schools that took part in our study aimed at integrating content and (second) language instruction by adopting a language sensitive instructional approach (e.g. Van Eerde & Hajer, 2005). Following this approach, language did not only play a central role in language classes, but also in classes whose overriding focus was on content (e.g. biology, mathematics, geography). The main aim of this language sensitive approach to content teaching is to cater for the special needs of L2 learners and learners with limited language proficiency, who might experience problems understanding and acquiring the content due to the linguistic demands of the input. The present investigation was conducted during biology classes, and our tasks treated biology-related topics.

The study's population consisted of seven intact classes of pupils in their second year of secondary pre-vocational education (or vmbo-t in Dutch) (N = 134), and six classes of pupils in their second year of higher general secondary education (or havo in Dutch) (N = 134). In the remainder of this chapter we will use the contrast 'lower-level' and 'higher-level' to refer to these two groups of pupils respectively. Pupils mean age was 14 (min. 14, max. 15). Within classes participants were randomly assigned to the different treatment groups incorporated in the study, to prevent a confounding interaction between treatment and
class. Furthermore, the male/female and Dutch L1/L2 proportion\(^3\) was kept constant across treatment groups.

### 4.5.2 Treatments

The present study integrated two experimental treatments and two control treatments:

**Experimental group I: direct CF (DIR)**  
Pupils in the first experimental group received comprehensive (or unfocused) direct CF. The researcher identified all existing linguistic errors, and provided the pupils with the corresponding target forms, as illustrated in example 1 below.

<table>
<thead>
<tr>
<th>Example 1: direct error correction</th>
</tr>
</thead>
<tbody>
<tr>
<td>De lieveheersbeestje is rood.</td>
</tr>
<tr>
<td>Het lieveheersbeestje</td>
</tr>
</tbody>
</table>

**Experimental group II: indirect CF (IND)**  
Texts of pupils in the second experimental group were corrected indirectly; the researcher provided an indication of each error and its category (cf. example 2), but it was left to the student to derive the corresponding target forms. We used nine different codes, classified under three coordinating categories, to correct different linguistic error types: (1) **lexical errors**: word choice; (2) **grammatical errors**: word form (e.g. verb tense, singular/plural), word order, incomplete sentences, and addition or omission of a word; and (3) **orthographical errors**: spelling, punctuation, and capitalization (cf. Appendix C).

---

\(^3\) We decided not to eliminate L1 speakers of Dutch from our sample because we consider ‘native-speakeriness’ to be a sliding scale. Since the large majority of our participants was born in the Netherlands, it was unfeasible to decide which pupils might have had access to (some) Dutch language input from birth onwards and which learners did not. What is shared by all our participants, however, is a relatively low level of writing proficiency (i.e. as judged by their teachers). This is why CF was expected to be beneficial to all learners in our sample, irrespective of their language background. Nevertheless, we did re-run all analyses excluding the pupils for whom Dutch was the only language spoken at home (N = 24). Results from these analyses did not differ in any respect from those reported in this chapter.
CHAPTER 4

Control group I: self-correction (SLF)

Pupils in the first control group did not receive any feedback on their initial text but were invited to revise their writing without any available CF – that is to self-correct their output. Including this control treatment enabled us to set apart effects of error correction from effects of the revision process as such.

Control group II: additional writing practice (PRC)

Just as the first control group, participants in the second control group did not receive CF. Whereas pupils in the first control group were invited to revise their own text, however, pupils in the second control group were not involved in any revision activities. Instead they performed a completely new writing task, which provided them with the opportunity to practice their writing skills once more. Inclusion of this second control treatment assured an equal distribution of time-on-task across experimental and control treatments, since completion of a completely new task asked (at least) as much time as revising an already written text. Moreover, integration of the practice treatment enabled us to determine if CF is more effective than just providing pupils with more practice opportunities (Truscott, 1996; 2004).

4.5.3 Procedure

The present investigation included four sessions; a pre-test session (S1), a treatment session (S2), a post-test session (S3), and a delayed post-test session (S4). Figure 4.1 presents an overview of the study’s set-up. Our tasks were designed for research 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. All tasks and tests were pen-and-paper assignments.

Example 2: indirect error correction

De lievenheersbeestje S is rood. [The ladybug is red.]

( _ = wrong word, S = spelling error)
Session 1: pre-tests (week 1)

During the first session (S1), all pupils (irrespective of their treatment group) were presented with a receptive vocabulary test (cf. Appendix E), a questionnaire concerning their language background (cf. Appendix D), and the first writing task. The output on this first assignment served as a baseline measure of pupils’ written accuracy and complexity. Pupils were given 20 minutes to complete the writing task, and were instructed to write a minimum of 15 lines. Scores on the vocabulary test were taken to provide an indication of pupils’ overall language proficiency. We included these scores in our analyses as a covariate to be able to control for individual L2 proficiency differences. A vocabulary test was chosen because earlier research demonstrated that vocabulary knowledge is a good predictor of overall language proficiency (e.g. Beglar & Hunt, 1999; Zareva, Schwanenflugel, & Nikolova, 2005). Moreover, a vocabulary test seemed to be the most suitable instrument out of practical considerations and time restrictions.

Before administering the first assignment, the researcher introduced the task’s topic by means of a ten minute mini-lesson, to ensure a comparable minimal amount of background knowledge on the topic amongst participants. Additionally, pupils were asked to focus on content as well as on form while writing. To give them an idea of the linguistic features they could attend to, pupils were given a hand-out listing common types of errors, and an example for each error category. We chose to draw learners’ attention to form at this stage, to be able to establish the unique contribution of CF in the end. If only the attention of the experimental groups would have been directed to linguistic form, it would have been impossible to unambiguously explain any apparent advantage for these groups over the control treatments; it would have been unclear if it was the form focus as such or the CF and/or revision that led to the effect. Moreover, the hand-out used here was re-used – in a slightly adapted version including error codes – when instructing pupils in the indirect CF group on interpreting the error codes (cf. section about session 2). This set-up ensured that the indirect group did not have an advantage as compared to the other groups in terms of language input. It is important to note that, whereas we directed pupils’ attention to language form, it was made clear that their main focus needed to be on the content of their writing. Furthermore, pupils were not aware of the fact that they could be receiving CF at a later stage.

Session 2: treatment (week 2)

The treatment session (S2) took place one week after S1. Pupils in the two CF groups received the corrected versions of their initial texts. They were asked to copy the initial
texts, revising all errors corrected by the researcher. Before starting their revision, pupils in the indirect CF group were given a hand-out (cf. Appendix C), accompanied by an oral instruction, on how to interpret and use the error codes in their texts. Pupils in the self-correction group were handed the texts they wrote during the pre-test session without any alterations. They were invited to self-correct their pre-test writing, by thoroughly reading over their texts and searching for any elements that needed revising. Even if no such utterances would be found, pupils were still obliged to copy their initial texts. (Cf. Appendix B for the instructions received by the direct CF, indirect CF, and self-correction groups.) Finally, pupils in the practice group were not given the opportunity to revise their pre-test texts, but were presented with a new writing task instead. The researcher shortly explained the new task’s topic before pupils started writing. All pupils were allocated 20 minutes to finish the task they were presented with, irrespective of the treatment group they were assigned to.

Sessions 3 and 4: post-test (week 3) and delayed post-test (week 6)

The first post-test (S3) was administered one week after the treatment session (S2), and the delayed post-test took place four weeks after S2. During both post-test sessions, all pupils were given 20 minutes to produce a text (at least 15 lines in length) on a new topic, which was again shortly introduced by the researcher.

Figure 4.1 Experimental set-up

<table>
<thead>
<tr>
<th>Treatment group</th>
<th>Pre-tests</th>
<th>Treatment</th>
<th>Post-test</th>
<th>Delayed post-test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>S1: week 1</td>
<td>S2: week 2</td>
<td>S3: week 3</td>
<td>S4: week 6</td>
</tr>
<tr>
<td>Direct CF (DIR)</td>
<td>Vocabulary test &amp; Language background questionnaire</td>
<td>revision based on comprehensive direct CF</td>
<td>New text</td>
<td>New text</td>
</tr>
<tr>
<td>Indirect CF (IND)</td>
<td>Initial text</td>
<td>revision based on comprehensive indirect CF</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-Correction (SLF)</td>
<td></td>
<td>revision based on self-correction (no CF)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Practice (PRC)</td>
<td>I</td>
<td>writing practice (no revision or CF)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
4.5.4 Writing tasks

The tasks used throughout the study were four writing assignments on a biology-related topic, namely the metamorphosis of different insects (i.e. butterflies (S1), honeybees (S2), ladybugs (S3), and wasps (S4)). Whereas three of the tasks were performed by all pupils, the task on a wasp's metamorphosis was only performed by participants in the second control group, who practiced their writing during the treatment session (S2).

All tasks within the series had a comparable form and set-up; they invited pupils to write an e-mail to a classmate who was absent during the researcher’s explanation of the task’s topic. Pupils were asked to explain the metamorphosis of the relevant insect, based on a series of images depicting the metamorphosis process. To avoid individual differences in familiarity with the relevant vocabulary, (potentially) unfamiliar words (e.g. larva, cocoon) were glossed. (Cf. Appendix A.)

4.5.5 Data processing

All handwritten output was transcribed and coded using CLAN (Computerized Language ANalysis) software, a program designed to analyze data transcribed in the CHILDES (Child Language Data Exchange System) format (MacWhinney, 2000). Two research assistants transcribed pupils’ texts, while the first author was responsible for coding them. Texts were coded for linguistic errors and clause types (i.e. main clauses and subordinate clauses). The coding procedure was entirely blind; during coding the researcher was unaware of the treatment group the text at hand belonged to. Ten percent of the data was also coded by one of the assistants to be able to establish inter-rater reliability, and re-coded, six months later, by the first author to measure intra-rater reliability.

We calculated intra-class correlation coefficients (ICC) to establish the average levels (over sessions) of intra-rater and inter-rater agreement for overall accuracy, grammatical accuracy, non-grammatical accuracy, and structural complexity (cf. section 4.5.6). The intra-rater ICC was calculated from an ANOVA two-way mixed effects model, and provides an indication of the variability due to variation within the same rater. A two-way random effects model was used to estimate the inter-rater ICC, reporting the proportion of the variability due to variation among raters. Results show high levels of agreement within the same rater as well as between raters for all four measures (cf. Table 4.1).

4 Since we only compared accuracy performance across groups and not sessions, tasks were not counterbalanced.
Table 4.1 Average levels of intra-rater and inter-rater agreement for measures of accuracy and structural complexity

<table>
<thead>
<tr>
<th></th>
<th>Overall Accuracy</th>
<th>Grammatical Accuracy</th>
<th>Non-grammatical Accuracy</th>
<th>Structural Complexity</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICC intra-rater</td>
<td>.998</td>
<td>.977</td>
<td>.996</td>
<td>.974</td>
</tr>
<tr>
<td>ICC inter-rater</td>
<td>.981</td>
<td>.921</td>
<td>.975</td>
<td>.946</td>
</tr>
</tbody>
</table>

4.5.6 Linguistic measures

Every text was analyzed for accuracy, structural complexity, and lexical diversity. As did earlier studies exploring the effectiveness of written CF (e.g. Chandler, 2003; Truscott & Hsu, 2008), we used an error ratio to measure overall accuracy (i.e. (number of form-related errors/total number of words) x 10). A ten-word ratio rather than the more common 100-word ratio was used because pupils’ texts were relatively short (i.e. around 120 words).

To be able to test Truscott’s (2001; 2007) claim that non-grammatical errors might be more correctable than errors in grammar, we broke down our overall accuracy measure into a measure of grammatical accuracy and a measure of non-grammatical accuracy. The first measure was a ratio calculated on the basis of the sum of the number of article errors, inflectional errors, word order errors, omissions of a necessary element, additions of a non-necessary element, pronominal errors, and other grammatical errors (i.e. (number of grammatical errors/total number of words) x 10). Lexical errors, orthographical errors, appropriateness/pragmatic errors, and other non-grammatical errors, on the other hand, were included in the non-grammatical accuracy ratio (i.e. (number of non-grammatical errors/total number of words) x 10).

To allow investigation of the possible influence of error correction on the complexity of pupils’ writing, we analyzed all texts for both structural complexity and lexical diversity. We used a subordination index, that is the number of subordinate clauses as a percentage of the total number of clauses (i.e. (number of sub clauses/total number of clauses) x 100), to obtain an indication of a text’s structural complexity (Norris & Ortega, 2009; Wolfe-

---

5 Although we were able to classify almost all of the observed errors, some did not fit in either of the specified error subcategories. These were brought together under ‘other grammatical errors’ or ‘other non-grammatical errors’.

6 The error categories we distinguished by different error codes when providing indirect CF (cf. Appendix C), are not in a one-to-one relation with the categories used during the coding procedure. For example, we used the same code – that is ‘wrong word’ – for article errors, pronominal errors, and lexical errors when providing indirect CF to keep the code labels as transparent as possible for our learners.
Lexical diversity was calculated using Guiraud’s Index, a type-token ratio that corrects for text length (types/√tokens) (Guiraud, 1954).

4.5.7 Analysis
As our participants came from different classes within different schools, our data were structured hierarchically. The performances of pupils within one class or school were expected to be more similar to each other than to performances of pupils in other classes and schools; a teacher, for example, is likely to affect the performances of his/her pupils. If observations within classes or schools had indeed been dependent, this would have affected the choice of appropriate statistical procedures. Analysis of variance for instance, relies on the independency of cases. We applied linear multilevel analyses (e.g. Snijders & Bosker, 1999) to explore the relationship between observations within classes and schools. The multilevel procedure enabled us to explicitly model possible dependencies in the data by including class and school as random factors. These multilevel analyses showed, however, that class and school never made a significant contribution to the model. After we had ascertained that the assumption of independence was met, we proceeded analyzing our data using AN(C)OVA’s because this procedure has proven to have more power in absence of data dependencies (e.g. Snijders & Bosker, 1999).

We used ANCOVA’s to test for between-group differences on the different dependent variables (i.e. overall accuracy, grammatical accuracy, non-grammatical accuracy, lexical diversity, and structural complexity) in the treatment session (S2), first post-test session (S3), and delayed post-test session (S4). The initial ANCOVA models contained treatment and educational level as between-subject variables, and language proficiency as a covariate. In addition, we incorporated learners’ pre-test (S1) performance on the relevant dependent variable as a covariate to account for effects of initial individual differences. We started out each ANCOVA with a full model, including all relevant factors, and all possible interactions between those factors. (Interactions between) factors, which did not explain a significant proportion of the variance, were excluded from the final ANCOVA models.

Because we only have two to four observations (i.e. classes) per school in our sample, we never included both random factors class and school in the same analysis. Instead we performed two separate analyses to find out if any of those contextual variables brought about unplanned dependencies in the data.
4.6 Results

In this section we will first present some relevant descriptive statistics, and describe learners’ performance on the writing and language proficiency pre-tests. We will then successively present the findings regarding the value of comprehensive CF as an editing tool (RQ 1), its language learning potential (RQ 2), the effect of direct and indirect CF on grammatical and non-grammatical issues (RQ’s 4 and 5), and the influence of CF on the complexity of learners’ writing (RQ 7). The comparison between the effects of CF and those of self-correction (RQ 3) and additional writing practice (RQ 7), and the role of pupils’ educational level (RQ 8), will be discussed throughout the results section.

4.6.1 Descriptive statistics

Language proficiency

The descriptive statistics for overall language proficiency (i.e. the score on a receptive vocabulary test), itemized per educational level and treatment group are presented in Table 4.2.

Table 4.2 Descriptive statistics: overall language proficiency\(^a\) by educational level and treatment

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Language proficiency</th>
<th>Treatment</th>
<th>Language proficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M  SD</td>
<td></td>
<td>M  SD</td>
</tr>
<tr>
<td>DIR (N = 32)</td>
<td>76.09 10.52</td>
<td>DIR (N = 34)</td>
<td>72.24 11.08</td>
</tr>
<tr>
<td>IND (N = 26)</td>
<td>74.85 11.80</td>
<td>IND (N = 31)</td>
<td>67.97 10.71</td>
</tr>
<tr>
<td>SLF (N = 33)</td>
<td>80.45 8.93</td>
<td>SLF (N = 31)</td>
<td>70.61 11.93</td>
</tr>
<tr>
<td>PRC (N = 33)</td>
<td>74.09 8.46</td>
<td>PRC (N = 33)</td>
<td>71.97 8.54</td>
</tr>
<tr>
<td><strong>Total (N = 124)</strong></td>
<td><strong>76.46 10.10</strong></td>
<td><strong>Total (N = 129)</strong></td>
<td><strong>70.75 10.63</strong></td>
</tr>
</tbody>
</table>

\(^a\)Number of correct items (out of 108) on receptive vocabulary test.

Accuracy

Table 4.3 displays the descriptive statistics for overall accuracy scores for all treatment groups, itemized per educational level and session (i.e. pre-test, treatment session, post-
test, and delayed post-test). Tables 4.4 and 4.5 provide the descriptives for grammatical and non-grammatical accuracy respectively.

**Table 4.3** Descriptive statistics: overall accuracy\(^a\) by educational level, treatment, and session

<table>
<thead>
<tr>
<th>Educational level</th>
<th>Treatment</th>
<th>Pre-test (S1)</th>
<th>Treatment session (S2)</th>
<th>Post-test (S3)</th>
<th>Delayed post-test (S4)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
<td>M</td>
</tr>
<tr>
<td>High (havo)</td>
<td>DIR (N = 32)</td>
<td>1.52</td>
<td>0.75</td>
<td>0.37</td>
<td>0.29</td>
</tr>
<tr>
<td></td>
<td>IND (N = 29)</td>
<td>1.25</td>
<td>0.67</td>
<td>0.42</td>
<td>0.32</td>
</tr>
<tr>
<td></td>
<td>SLF (N = 37)</td>
<td>1.39</td>
<td>0.60</td>
<td>1.28</td>
<td>0.66</td>
</tr>
<tr>
<td></td>
<td>PRC (N = 36)</td>
<td>1.29</td>
<td>0.63</td>
<td>1.30</td>
<td>0.72</td>
</tr>
<tr>
<td></td>
<td>Total (N = 134)</td>
<td>1.37</td>
<td>0.66</td>
<td>0.88</td>
<td>0.71</td>
</tr>
<tr>
<td>Low (vmbo-t)</td>
<td>DIR (N = 35)</td>
<td>1.69</td>
<td>0.70</td>
<td>0.32</td>
<td>0.24</td>
</tr>
<tr>
<td></td>
<td>IND (N = 33)</td>
<td>1.95</td>
<td>0.88</td>
<td>0.74</td>
<td>0.63</td>
</tr>
<tr>
<td></td>
<td>SLF (N = 34)</td>
<td>1.85</td>
<td>0.75</td>
<td>1.50</td>
<td>0.68</td>
</tr>
<tr>
<td></td>
<td>PRC (N = 32)</td>
<td>1.79</td>
<td>1.00</td>
<td>2.19</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>Total (N = 134)</td>
<td>1.82</td>
<td>0.83</td>
<td>1.17</td>
<td>0.99</td>
</tr>
</tbody>
</table>

\(^a\)Number of form-related errors per 10 words.

**Table 4.4** Descriptive statistics: grammatical accuracy\(^a\) by educational level, treatment, and session

<table>
<thead>
<tr>
<th>Educational level</th>
<th>Treatment</th>
<th>Pre-test (S1)</th>
<th>Treatment session (S2)</th>
<th>Post-test (S3)</th>
<th>Delayed post-test (S4)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
<td>M</td>
</tr>
<tr>
<td>High (havo)</td>
<td>DIR (N = 32)</td>
<td>0.43</td>
<td>0.27</td>
<td>0.09</td>
<td>0.10</td>
</tr>
<tr>
<td></td>
<td>IND (N = 29)</td>
<td>0.30</td>
<td>0.24</td>
<td>0.10</td>
<td>0.11</td>
</tr>
<tr>
<td></td>
<td>SLF (N = 37)</td>
<td>0.37</td>
<td>0.25</td>
<td>0.37</td>
<td>0.23</td>
</tr>
<tr>
<td></td>
<td>PRC (N = 36)</td>
<td>0.37</td>
<td>0.27</td>
<td>0.36</td>
<td>0.27</td>
</tr>
<tr>
<td></td>
<td>Total (N = 134)</td>
<td>0.37</td>
<td>0.26</td>
<td>0.24</td>
<td>0.24</td>
</tr>
<tr>
<td>Low (vmbo-t)</td>
<td>DIR (N = 35)</td>
<td>0.55</td>
<td>0.31</td>
<td>0.10</td>
<td>0.11</td>
</tr>
<tr>
<td></td>
<td>IND (N = 33)</td>
<td>0.47</td>
<td>0.33</td>
<td>0.21</td>
<td>0.23</td>
</tr>
<tr>
<td></td>
<td>SLF (N = 34)</td>
<td>0.53</td>
<td>0.32</td>
<td>0.46</td>
<td>0.29</td>
</tr>
<tr>
<td></td>
<td>PRC (N = 32)</td>
<td>0.52</td>
<td>0.36</td>
<td>0.59</td>
<td>0.41</td>
</tr>
<tr>
<td></td>
<td>Total (N = 134)</td>
<td>0.52</td>
<td>0.33</td>
<td>0.24</td>
<td>0.24</td>
</tr>
</tbody>
</table>

\(^a\)Number of grammatical errors per 10 words.
### Table 4.5 Descriptive statistics: non-grammatical accuracy\(^a\) by educational level, treatment, and session

<table>
<thead>
<tr>
<th>Educational level</th>
<th>Treatment</th>
<th>Pre-test (S1)</th>
<th>Treatment session (S2)</th>
<th>Post-test (S3)</th>
<th>Delayed post-test (S4)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
<td>M</td>
</tr>
<tr>
<td>High (havo)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DIR (N = 32)</td>
<td>1.09</td>
<td>0.63</td>
<td>0.28</td>
<td>0.27</td>
<td>0.77</td>
</tr>
<tr>
<td>IND (N = 29)</td>
<td>0.95</td>
<td>0.52</td>
<td>0.32</td>
<td>0.26</td>
<td>0.69</td>
</tr>
<tr>
<td>SLF (N = 37)</td>
<td>1.01</td>
<td>0.55</td>
<td>0.92</td>
<td>0.56</td>
<td>1.20</td>
</tr>
<tr>
<td>PRC (N = 36)</td>
<td>0.93</td>
<td>0.54</td>
<td>0.94</td>
<td>0.63</td>
<td>1.05</td>
</tr>
<tr>
<td><strong>Total</strong> (N = 134)</td>
<td>1.00</td>
<td>0.56</td>
<td>0.64</td>
<td>0.56</td>
<td>0.95</td>
</tr>
<tr>
<td>Low (vmbo-t)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DIR (N = 35)</td>
<td>1.14</td>
<td>0.56</td>
<td>0.23</td>
<td>0.17</td>
<td>1.15</td>
</tr>
<tr>
<td>IND (N = 33)</td>
<td>1.48</td>
<td>0.72</td>
<td>0.52</td>
<td>0.47</td>
<td>1.22</td>
</tr>
<tr>
<td>SLF (N = 34)</td>
<td>1.32</td>
<td>0.63</td>
<td>1.05</td>
<td>0.64</td>
<td>1.33</td>
</tr>
<tr>
<td>PRC (N = 32)</td>
<td>1.28</td>
<td>0.82</td>
<td>1.60</td>
<td>0.85</td>
<td>1.56</td>
</tr>
<tr>
<td><strong>Total</strong> (N = 134)</td>
<td>1.30</td>
<td>0.69</td>
<td>0.83</td>
<td>0.78</td>
<td>1.32</td>
</tr>
</tbody>
</table>

\(^a\)Number of non-grammatical errors per 10 words.

Figure 4.2 is a graphic illustration of the descriptives presented in Table 4.3 (collapsed over educational level). It shows the accuracy development of the different treatment groups over time. The graph provides a descriptive preview of this study’s main findings, which will be presented in the following section. As can be seen from the graph, all groups performed similar at the pre-test (S1). In the treatment session (S2), however, the error rate of the CF groups and the self-correction group decreased, whereas the number of errors committed by the practice group increased. In the two post-tests (S3 and S4), we still see an error rate difference between the control and experimental groups, in favor of the latter.
Figure 4.2 Overall accuracy per treatment and session
CHAPTER 4

Complexity
Table 4.6 shows the descriptives for structural complexity per treatment group, itemized per educational level and session. Lexical diversity scores are presented in Table 4.7.

Table 4.6 Descriptive statistics: structural complexity by educational level, treatment, and session

<table>
<thead>
<tr>
<th>Educational level</th>
<th>Treatment</th>
<th>Pre-test (S1)</th>
<th>Treatment session (S2)</th>
<th>Post-test (S3)</th>
<th>Delayed post-test (S4)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
<td>M</td>
</tr>
<tr>
<td>High (havo)</td>
<td>DIR (N = 32)</td>
<td>16.31</td>
<td>10.46</td>
<td>17.30</td>
<td>10.87</td>
</tr>
<tr>
<td></td>
<td>IND (N = 29)</td>
<td>16.30</td>
<td>8.98</td>
<td>18.07</td>
<td>10.20</td>
</tr>
<tr>
<td></td>
<td>SLF (N = 37)</td>
<td>14.74</td>
<td>8.94</td>
<td>14.39</td>
<td>9.30</td>
</tr>
<tr>
<td></td>
<td>PRC (N = 36)</td>
<td>19.86</td>
<td>10.02</td>
<td>16.13</td>
<td>8.73</td>
</tr>
<tr>
<td>Total (N = 134)</td>
<td>16.83</td>
<td>9.71</td>
<td>16.35</td>
<td>9.74</td>
<td>15.16</td>
</tr>
</tbody>
</table>

Low (vmbo-t)

<table>
<thead>
<tr>
<th>Educational level</th>
<th>Treatment</th>
<th>Pre-test (S1)</th>
<th>Treatment session (S2)</th>
<th>Post-test (S3)</th>
<th>Delayed post-test (S4)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
<td>M</td>
</tr>
<tr>
<td>Low (vmbo-t)</td>
<td>DIR (N = 35)</td>
<td>20.78</td>
<td>8.17</td>
<td>21.66</td>
<td>8.27</td>
</tr>
<tr>
<td></td>
<td>IND (N = 33)</td>
<td>19.41</td>
<td>11.66</td>
<td>21.04</td>
<td>10.57</td>
</tr>
<tr>
<td></td>
<td>SLF (N = 34)</td>
<td>15.03</td>
<td>8.95</td>
<td>16.19</td>
<td>8.62</td>
</tr>
<tr>
<td></td>
<td>PRC (N = 32)</td>
<td>18.66</td>
<td>8.00</td>
<td>15.19</td>
<td>8.23</td>
</tr>
<tr>
<td>Total (N = 134)</td>
<td>18.52</td>
<td>9.45</td>
<td>18.57</td>
<td>9.32</td>
<td>15.89</td>
</tr>
</tbody>
</table>

\*Subordination Index (i.e. (number of sub clauses/total number of clauses) x 100)

Table 4.7 Descriptive statistics: lexical diversity by educational level, treatment, and session

<table>
<thead>
<tr>
<th>Educational level</th>
<th>Treatment</th>
<th>Pre-test (S1)</th>
<th>Treatment session (S2)</th>
<th>Post-test (S3)</th>
<th>Delayed post-test (S4)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
<td>M</td>
</tr>
<tr>
<td>High (havo)</td>
<td>DIR (N = 32)</td>
<td>7.13</td>
<td>0.85</td>
<td>7.28</td>
<td>0.91</td>
</tr>
<tr>
<td></td>
<td>IND (N = 29)</td>
<td>7.10</td>
<td>0.77</td>
<td>7.10</td>
<td>0.73</td>
</tr>
<tr>
<td></td>
<td>SLF (N = 37)</td>
<td>7.09</td>
<td>0.81</td>
<td>7.06</td>
<td>0.78</td>
</tr>
<tr>
<td></td>
<td>PRC (N = 36)</td>
<td>7.31</td>
<td>0.50</td>
<td>6.72</td>
<td>0.72</td>
</tr>
<tr>
<td>Total (N = 134)</td>
<td>7.16</td>
<td>0.74</td>
<td>7.03</td>
<td>0.81</td>
<td>6.79</td>
</tr>
</tbody>
</table>

Low (vmbo-t)

<table>
<thead>
<tr>
<th>Educational level</th>
<th>Treatment</th>
<th>Pre-test (S1)</th>
<th>Treatment session (S2)</th>
<th>Post-test (S3)</th>
<th>Delayed post-test (S4)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
<td>M</td>
</tr>
<tr>
<td>Low (vmbo-t)</td>
<td>DIR (N = 35)</td>
<td>7.00</td>
<td>0.81</td>
<td>7.26</td>
<td>0.70</td>
</tr>
<tr>
<td></td>
<td>IND (N = 33)</td>
<td>6.77</td>
<td>1.06</td>
<td>7.03</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>SLF (N = 34)</td>
<td>7.08</td>
<td>0.75</td>
<td>7.12</td>
<td>0.82</td>
</tr>
<tr>
<td></td>
<td>PRC (N = 32)</td>
<td>6.96</td>
<td>0.87</td>
<td>6.53</td>
<td>0.77</td>
</tr>
<tr>
<td>Total (N = 134)</td>
<td>6.95</td>
<td>0.88</td>
<td>7.00</td>
<td>0.87</td>
<td>6.55</td>
</tr>
</tbody>
</table>

\*Guiraud’s Index (types/√tokens)
4.6.2 Pre-test performance

A series of ANOVA's on pupils' pre-test texts (S1) showed that there were no initial differences between treatment groups in overall accuracy \((F (3, 264) < 1, p = .902, \eta_p^2 = .00)\), grammatical accuracy \((F (3, 264) = 1.30, p = .276, \eta_p^2 = .02)\), non-grammatical accuracy \((F (3, 264) < 1, p = .620, \eta_p^2 = .01)\), and lexical diversity \((F (3, 264) < 1, p = .484, \eta_p^2 = .01)\). We did find a significant initial difference on our measure of structural complexity \((F (3, 264) = 3.04, p = .029, \eta_p^2 = .03)\). Post-hoc pair wise comparisons revealed that the pre-test writing of pupils in the practice group was structurally more complex than that of learners in the self-correction group \((p = .033)\). Finally, we did not find a significant difference between treatment groups in language proficiency \((F (3, 253) = 1.96, p = .120, \eta_p^2 = .02)\).

Educational level turned out to be a significant predictor of language proficiency \((F (1, 251) = 19.15, p < .001, \eta_p^2 = .07)\); pupils with a higher educational level scored higher on the vocabulary test than pupils with a lower level of education.

Moreover, results did display a significant effect of educational level and language proficiency on learners' pre-test writing. We found that both educational level \((F (1, 250) = 14.20, p < .001, \eta_p^2 = .05)\) and language proficiency \((F (1, 250) = 15.45, p < .001, \eta_p^2 = .06)\) predicted the overall accuracy of pupils' S1 texts in such a way that the lower-level pupils committed significantly more errors than pupils with a higher level of education and language proficiency. The same holds for grammatical accuracy (educational level: \(F (1, 250) = 10.83, p = .001, \eta_p^2 = .04\); language proficiency: \(F (1, 250) = 12.69, p < .001, \eta_p^2 = .05\)) and non-grammatical accuracy (educational level: \(F (1, 250) = 8.81, p = .003, \eta_p^2 = .03\); language proficiency: \(F (1, 250) = 9.21, p = .003, \eta_p^2 = .04\)). Finally, the lexical diversity (but not the structural complexity) of pupils' pre-test writing co-varied significantly with their language proficiency \((F (1, 250) = 7.98, p = .005, \eta_p^2 = .03)\).

4.6.3 Effects of comprehensive CF on written accuracy

As mentioned in section 4.5.7, ANCOVA's were used to test for between-group accuracy differences in the treatment session (S2), post-test (S3), and delayed post-test (S4). The initial ANCOVA models contained treatment and educational level as between-subject variables, and language proficiency and overall accuracy S1 as covariates. Insignificant (interactions between) factors were excluded from the final models.
Revision effects (S2)
During revision, pupils’ language proficiency ($F(1, 240) = 2.63, p = .106, \eta^2_p = .01$) and the interaction between treatment and language proficiency ($F(3, 240) = 2.40, p = .069, \eta^2_p = .03$) did not prove to play a significant role. These factors were therefore discarded from the final ANCOVA model. The definitive analysis revealed significant effects of treatment ($F(3, 259) = 111.24, p < .001, \eta^2_p = .56$) and educational level ($F(1, 259) = 4.29, p = .039, \eta^2_p = .02$). We also found that the interaction between treatment and educational level made a significant contribution to the model ($F(3, 259) = 7.85, p < .001, \eta^2_p = .08$). Furthermore, accuracy in the treatment session (S2) co-varied significantly with pre-test accuracy scores (S1) ($F(1, 259) = 112.77, p < .001, \eta^2_p = .30$).

When comparing the sizes of the different significant effects, it is interesting to note that the treatment learners received and their performance on the pre-test text (S1), have a large effect on the treatment session (S2) outcomes. Their educational level, on the other hand, only rendered a medium sized effect.

To be able to determine where the significant differences in overall accuracy lay between treatment groups, we conducted post-hoc pairwise comparisons, using Bonferroni adjustments. Table 4.8 presents an overview of all significant contrasts between treatments per session and accuracy type. The most important observation is that pupils were able to use the CF they received during revision; both CF treatments turned out to be significantly more beneficial than either of the control treatments. Pupils who received direct CF outperformed pupils in the self-correction ($p < .001$, Cohen’s $d = 1.99$) and practice group ($p < .001$, Cohen’s $d = 2.74$). Indirect CF also proved to be more constructive than both correcting without feedback ($p < .001$, Cohen’s $d = 1.55$) and practicing writing ($p < .001$, Cohen’s $d = 2.29$). A comparison between the two experimental groups did not reach significance. However, the Cohen’s $d$ values show that the effects of direct CF were larger than those of indirect CF. This was to be expected, because pupils receiving direct CF only needed to copy the target forms as provided by the researcher. We furthermore did observe significant differences between the two control treatments, which can be explained from their interaction with pupils’ educational level. Whereas the higher-level pupils benefited equally from practicing their writing and self-correcting their texts, self-correction had an advantage over writing practice for the lower-level pupils ($p < .001$, Cohen’s $d = 1.23$) during the treatment session.

---

8 We applied the following rule of thumb in interpreting $\eta^2_p$ values: small $0.01 \leq \eta^2_p < 0.06$, medium $0.06 \leq \eta^2_p < 0.14$, large $\eta^2_p \geq 0.14$.

9 According to Cohen (1988), effect sizes can be classified as either small (0.20), medium (0.50), or large (0.80).
Table 4.8 Summary of significant contrasts between treatment groups

<table>
<thead>
<tr>
<th></th>
<th>Overall accuracy</th>
<th>Grammatical accuracy</th>
<th>Non-grammatical accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment session (S2)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>***DIR &gt; SLF</td>
<td>***DIR &gt; SLF</td>
<td>***DIR &gt; SLF</td>
<td></td>
</tr>
<tr>
<td>***DIR &gt; PRC</td>
<td>***DIR &gt; PRC</td>
<td>***DIR &gt; PRC</td>
<td></td>
</tr>
<tr>
<td>***IND &gt; SLF</td>
<td>***IND &gt; SLF</td>
<td>***IND &gt; SLF</td>
<td></td>
</tr>
<tr>
<td>***IND &gt; PRC</td>
<td>***IND &gt; PRC</td>
<td>***IND &gt; PRC</td>
<td></td>
</tr>
<tr>
<td>Post-test (S3)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>***DIR &gt; SLF</td>
<td>*DIR &gt; PRC</td>
<td>*DIR &gt; SLF</td>
<td></td>
</tr>
<tr>
<td>***DIR &gt; PRC</td>
<td>**DIR &gt; PRC</td>
<td>**DIR &gt; PRC</td>
<td></td>
</tr>
<tr>
<td>***IND &gt; SLF</td>
<td>**IND &gt; SLF</td>
<td>**IND &gt; SLF</td>
<td></td>
</tr>
<tr>
<td>***IND &gt; PRC</td>
<td>**IND &gt; PRC</td>
<td>**IND &gt; PRC</td>
<td></td>
</tr>
<tr>
<td>Delayed post-test (S4)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>**DIR &gt; PRC</td>
<td>**DIR &gt; PRC</td>
<td>**IND &gt; SLF</td>
<td></td>
</tr>
<tr>
<td>**IND &gt; SLF</td>
<td>**IND &gt; SLF</td>
<td>**IND &gt; PRC</td>
<td></td>
</tr>
<tr>
<td>***IND &gt; PRC</td>
<td>**IND &gt; PRC</td>
<td>**IND &gt; PRC</td>
<td></td>
</tr>
</tbody>
</table>

*p<.05, **p<.01, ***p<.001.

Short-term learning effects (S3)

The initial ANCOVA model applied to the data of our first post-test (i.e. the texts written one week after pupils received the different treatments) revealed that the interactions between treatment and educational level ($F(3, 225) < 1, p = .519, \eta_p^2 = .01$), and treatment and language proficiency ($F(3, 225) < 1, p = .436, \eta_p^2 = .01$) were not significant. These factors were therefore not included in the definitive ANCOVA model. Our final analysis showed that the four treatment groups significantly varied in the number of errors they made in the first post-test ($F(3, 231) = 11.34, p < .001, \eta_p^2 = .13$). Again, we found that pupils with a higher level of education outperformed those with a lower educational level ($F(1, 231) = 4.45, p = .036, \eta_p^2 = .02$). Moreover, S3 accuracy scores significantly co-varied with both language proficiency ($F(1, 231) = 12.93, p < .001, \eta_p^2 = .05$), and overall accuracy S1 ($F(1, 231) = 87.90, p < .001, \eta_p^2 = .28$). As was the case in the treatment session (S2), especially the effects of treatment and pre-test performance are noteworthy. Compared to the effects of these two factors, the effects associated with learners’ educational level and language proficiency are relatively small.

In the light of the treatment effect that was found, we performed post-hoc pair wise comparisons, using Bonferroni adjustments (cf. Table 4.8). We observed that both CF groups outperformed the two control groups. Pupils who received direct CF on the first piece of writing, made fewer errors in the new text they wrote a week later (S3), than pupils who self-corrected their errors during revision ($p = .002, \text{Cohen’s } d = 0.67$) or practiced their writing skills without any revising opportunity ($p < .001, \text{Cohen’s } d = 0.81$). Pupils whose
errors were corrected indirectly also performed significantly better on the first post-test than pupils in the self-correction ($p = .001$, Cohen’s $d = 0.70$) and practice ($p < .001$, Cohen’s $d = 0.84$) groups. There were no significant differences amongst the CF treatments; direct and indirect CF showed to be equally effective in improving pupils’ written accuracy (when the new text was written a week after the moment of feedback provision). When comparing the sizes of the S3 CF effects to the Cohen’s $d$ values reported in the previous section, we see that CF has a smaller effect on the accuracy of new texts than on revisions. To conclude, no significant accuracy differences were found between the two control groups.

**Long-term learning effects (S4)**

As in the first post-test session (S3), the interactions between treatment and educational level ($F(3, 205) < 1$, $p = .649$, $\eta^2_p = .01$), and treatment and language proficiency ($F(3, 205) < 1$, $p = .952$, $\eta^2_p = .00$), turned out to be non-significant during the delayed post-test (S4), which took place four weeks after the treatment session. Moreover, pupils’ educational level was no longer an explaining factor ($F(1, 205) < 1$, $p = .436$, $\eta^2_p = .00$). These non-significant parameters were deleted from the ANCOVA model. The final model revealed that there was a significant between-group difference concerning overall accuracy in the texts learners wrote four weeks after they received CF, self-corrected their initial text, or practiced their writing skills ($F(3, 212) = 8.45$, $p < .001$, $\eta^2_p = .11$). As on the first post-test, language proficiency ($F(1, 212) = 5.56$, $p = .019$, $\eta^2_p = .02$) and overall accuracy S1 ($F(1, 212) = 86.38$, $p < .001$, $\eta^2_p = .29$) proved to be significant covariates. Once more, the effect of learners’ language proficiency was much smaller than those of treatment and pupils’ performance on the pre-test (S1).

Bonferroni adjusted post-hoc pair wise comparisons revealed that pupils whose errors were corrected four weeks before writing a new text, still benefited from the CF they received (cf. Table 4.8). Both the direct ($p = .005$, Cohen’s $d = 0.63$) and indirect ($p < .001$, Cohen’s $d = 0.94$) groups made significantly fewer errors than pupils who practiced their writing skills during the treatment session (S2). Table 4.8 furthermore shows that pupils who received indirect CF also outperformed the group who corrected their own errors without any available feedback ($p = .005$, Cohen’s $d = 0.66$). Interestingly, the post-test (S3) and delayed post-test (S4) effects are comparable in size. This suggests that the CF invoked accuracy gains are not transient in nature.
4.6.4 Effects of direct and indirect CF on grammatical and non-grammatical errors

To test Truscott’s (2007) claim that CF could have value for non-grammatical errors but not for errors in grammar, we performed separate analyses for grammatical and non-grammatical error types. The treatment session (S2), post-test (S3), and delayed post-test (S4) outcomes of two series of ANCOVA’s are reported; one for grammatical accuracy and one for non-grammatical accuracy. The initial ANCOVA models contained treatment and educational level as between-subject variables, and language proficiency and (non-)grammatical accuracy S1 as covariates. To increase the clarity of our findings, this section only reports the observed significant (interactions between) main effects.

Revision effects (S2)

During revision, the treatment learners received proved to have a significant effect on both pupils’ grammatical accuracy \((F(3, 259) = 51.09, p < .001, \eta^2_p = .37)\) and non-grammatical accuracy \((F(3, 259) = 81.80, p < .001, \eta^2_p = .49)\). The outcomes of post-hoc pair wise comparisons with Bonferroni adjustments per accuracy type, showed that CF enabled learners to resolve both their grammatical and non-grammatical errors when revising their initial text (cf. Table 4.8). Pupils who received direct CF outperformed learners in the self-correction group on both grammatical \((p < .001, \text{Cohen’s } d = 1.52)\) and non-grammatical \((p < .001, \text{Cohen’s } d = 1.61)\) accuracy. Moreover, the direct group made significantly fewer grammatical \((p < .001, \text{Cohen’s } d = 1.85)\) and non-grammatical \((p < .001, \text{Cohen’s } d = 2.35)\) errors than pupils who practiced their writing skills during the treatment session (S2). Indirect CF also proved to have a more beneficial effect on both accuracy measures than either self-correction (grammatical accuracy: \(p < .001, \text{Cohen’s } d = 1.05\); non-grammatical accuracy: \(p < .001, \text{Cohen’s } d = 1.33\)) or writing practice (grammatical accuracy: \(p < .001, \text{Cohen’s } d = 1.37\); non-grammatical accuracy: \(p < .001, \text{Cohen’s } d = 2.06\)). The effects associated with direct CF were again greater than those of indirect CF. Analyses furthermore revealed that the interaction we found on our measure of overall accuracy – between the effectiveness of the different control treatments and pupils’ educational level – was related to non-grammatical accuracy only. The self-correction and practice treatments were equally (in)effective for pupils of both educational levels with respect to grammatical accuracy. However, the influence of the two control treatments on non-grammatical accuracy differed significantly across educational levels. Whereas the practice and self-correction treatments were equally effective for the higher-level pupils, the lower-level
learners benefited more from self-correcting their non-grammatical errors than from writing practice ($p < .001$, Cohen’s $d = 1.15$).

**Short-term learning effects (S3)**

ANCOVA’s revealed that there was a significant difference between groups one week after the treatment session (i.e. during S3) in performance on both the grammatical ($F(3, 231) = 4.10$, $p = .007$, $\eta^2_p = .05$) and non-grammatical ($F(3, 232) = 8.95$, $p < .001$, $\eta^2_p = .10$) accuracy measures. Post-hoc pair wise comparisons showed, however, that the effectiveness of direct and indirect CF differed across accuracy types (cf. Table 4.8). Both CF approaches were more effective in improving pupils’ non-grammatical accuracy than self-correction (direct CF: $p = .021$, Cohen’s $d = 0.53$; indirect CF: $p = .002$, Cohen’s $d = 0.68$) or writing practice (direct CF: $p = .003$, Cohen’s $d = 0.65$; indirect CF: $p < .001$, Cohen’s $d = 0.80$). In contrast, only direct CF helped pupils to reduce the number of grammatical errors in a new text which was written one week after the feedback had been provided; direct CF proved to be significantly more beneficial than practicing writing ($p = .016$, Cohen’s $d = 0.55$).

**Long-term learning effects (S4)**

When we considered pupils’ grammatical and non-grammatical performance on the delayed post-test (S4), we still found a significant between-group difference on both accuracy measures (grammatical: $F(3, 212) = 4.71$, $p = .003$, $\eta^2_p = .06$, non-grammatical: $F(3, 212) = 6.60$, $p < .001$, $\eta^2_p = .08$). Post-hoc pair wise comparisons revealed that the positive effect of direct CF on grammatical accuracy was still present four weeks after the treatment had taken place (cf. Table 4.8); pupils in the direct group made fewer grammatical errors than classmates who were allocated additional practice opportunity ($p = .004$, Cohen’s $d = 0.65$). Furthermore, whereas pupils’ non-grammatical accuracy benefited from both direct and indirect CF during the first post-test (S3), only the effect of indirect CF was sustained in the delayed post-test. Results showed indirect CF to be significantly more effective in reducing the number of non-grammatical errors in pupils’ writing than either self-correction ($p = .003$, Cohen’s $d = 0.66$) or writing practice ($p < .001$, Cohen’s $d = 0.82$).

**4.6.5 Effects of CF on written complexity**

We did not only aim at investigating the effect of CF on written accuracy, but also explored the influence of error correction on the structural complexity and lexical diversity of pupils’ writing. In doing so, we tested Truscott’s (2007) hypothesis that CF might lead to avoidance
of more complex structures. The initial ANCOVA models reported in this section contained treatment and educational level as between-subject variables, and language proficiency and either lexical or structural complexity S1 as covariates. Educational level and language proficiency were excluded from the final model since these factors turned out to be insignificant.

Our data did not confirm Truscott’s assumption that CF leads to simplified writing; we did not find any significant between-group differences on our measures of lexical diversity or structural complexity in either of the post-tests (structural complexity S3: \( F (3, 244) = 1.45, \ p = .229, \ \eta^2_p = .02 \); lexical diversity S3: \( F (3, 244) < 1, \ p = .712, \ \eta^2_p = .01 \); structural complexity S4: \( F (3, 227) = 1.31, \ p = .271, \ \eta^2_p = .02 \); lexical diversity S4: \( F (3, 227) = 1.85, \ p = .138, \ \eta^2_p = .02 \)). Results did reveal significant between-group differences concerning the structural complexity (\( F (3, 262) = 7.94, \ p < .001, \ \eta^2_p = .08 \)) and lexical diversity (\( F (3, 262) = 31.65, \ p < .001, \ \eta^2_p = .27 \)) of the output produced during the treatment session (S2). The writing of pupils who wrote a new text during S2 (i.e. the practice group) was structurally less complex than the writing of pupils who received direct CF (\( p = .001, \ \text{Cohen’s} \ d = .68 \)) or indirect CF (\( p < .001, \ \text{Cohen’s} \ d = .77 \)). Furthermore, all pupils who had the opportunity to revise their text, outperformed the practice group on the measure of lexical diversity (direct CF: \( p < .001, \ \text{Cohen’s} \ d 1.55 \); indirect CF: \( p < .001, \ \text{Cohen’s} \ d 1.32 \); self-correction: \( p < .001, \ \text{Cohen’s} \ d 1.11 \)).

4.7 General Discussion

Truscott (1996; 1999; 2004; 2007; 2009) has repeatedly argued that CF has no place in L2 classrooms because grammar correction would be more likely to hinder than to facilitate accuracy development. Our findings, however, clearly show that comprehensive CF is an effective means of improving learners’ accuracy over time, and do not support Truscott’s supposition that CF has detrimental side-effects.

4.7.1 The language learning potential of comprehensive CF

The main aim of the present study was to investigate whether **comprehensive or unfocused** CF leads to improved accuracy in L2 writing. Just as earlier studies (e.g. Ashwell, 2000; Fathman & Whalley, 1990; Ferris, 1997; Ferris & Roberts, 2001), we found that comprehensive CF enables learners to enhance the linguistic correctness of a certain manuscript during revision (RQ 1). The important contribution our study makes to the
existing literature, however, relates to the effectiveness of comprehensive or unfocused CF in expanding pupils’ accuracy in new texts (RQ 2). Whereas research already revealed that learners receiving focused CF are able to use the targeted form more accurately in new pieces of writing (e.g. Bitchener, 2008; Bitchener & Knoch, 2008; Bitchener & Knoch, 2009; Bitchener & Knoch, 2010a; Bitchener & Knoch, 2010b; Ellis et al., 2008; Sheen, 2007; Sheen, 2010b), our study is the first to show that unfocused CF leads to learning; we found that pupils whose errors were corrected comprehensively made fewer errors in new pieces of writing than learners who did not receive CF. The fact that receiving CF proved to be more beneficial than self-correction without any available feedback, furthermore shows that CF has an added value above revision as such (RQ 3). Importantly, the positive effects of comprehensive CF showed to be durable; accuracy gains were visible both in the post-test and the delayed post-test, which respectively took place one and four weeks after pupils received the corrections. (Cf. Table 4.8.) Our post-test results indicated medium effect sizes (on average) for the advantage of CF over practicing writing and self-correction. This finding concurs with those of meta-analyses investigating the effectiveness of CF (Li, 2010; Lyster & Saito, 2010; Russell & Spada, 2006), which consistently reported medium effects\(^\text{10}\) of error correction on immediate and delayed post-tests, in both oral and written settings, and across laboratory and classroom-based studies.

4.7.2 Interactions between feedback methodology and error type

Apart from exploring the overall effectiveness of comprehensive CF, the present study investigated the relative efficacy of direct and indirect correction methodologies (RQ 4), and the differential effects of CF on grammatical and non-grammatical error types (RQ 5). We found that different CF types have value for different types of errors; whereas only direct correction promoted grammatical accuracy improvement, pupils’ non-grammatical accuracy benefited most from indirect CF (i.e. the effect of indirect CF was retained the longest). (Cf. Table 4.8.) These findings are in line with suggestions in the literature concerning the relative effectiveness of direct and indirect CF. The cognitively demanding indirect approach did foster long-term acquisition of those linguistic features pupils were able and confident to self-correct (using the error codes provided), that is their non-grammatical errors (Bitchener

\(^{10}\) Whereas Lyster and Saito (2010) and Russell and Spada (2006) reported large overall effect sizes for CF, both studies concluded that CF has a medium effect on L2 acquisition; Lyster and Saito inferred that the effects of CF could be considered medium by subtracting the effect found for the control groups from the overall large effect of CF. When reconsidering the methodological quality of the studies included in their meta-analysis, Russell and Spada noted that studies reporting reliable and valid measures yielded medium effect sizes.
& Knoch, 2008; Ferris, 1995). Conversely, although indirect CF did supply learners with sufficient information to resolve their more complex, grammatical errors during revision, the fact that they did not know if their own hypothesized corrections were accurate might have prevented them from internalizing the correct structures (cf. Chapter 3; Chandler, 2003). Only when the grammatical errors were corrected directly, pupils showed evidence of long-term learning of the corresponding target forms. Based on these findings, we conclude that both grammatical and non-grammatical errors are amenable to CF, but that they benefit from different types of correction. This leads us to suggest an alternative explanation for the apparent lack of a learning effect in Truscott and Hsu's (2008) study. The error categories targeted in their study were mainly grammatical in nature; orthographical and lexical errors were not corrected. Moreover, an indirect type of CF (i.e. underlining of errors) was used. Our results indicate, however, that only direct CF has the potential to yield long-term grammatical accuracy gains.

4.7.3 The potential detrimental side-effects of CF

We also aimed at testing Truscott's (1996; 2004; 2007) hypothesis that CF could be harmful to learners' accuracy development, because (i) CF diverts time and energy away from more productive aspects of writing instruction, such as additional writing practice (RQ 6), and (ii) CF has a negative influence on the complexity of L2 learners' writing, in that it makes them avoid structures that have been corrected before (RQ 7). Our findings oppose both of these claims. CF did not lead our participants to produce lexically or structurally less complex writing, and the texts written by pupils who received CF were more accurate than those of learners who were allowed an extra opportunity to practice their writing skills.

4.7.4 Educational level and CF efficacy

A final issue the present study explored, was the possible influence of participants’ educational level on feedback effectiveness (RQ 8). The main reason for exploring this issue was practical in nature. However, it is also theoretically plausible that educational level mediates the efficacy of indirect CF. We presumed the higher-level pupils to dispose over more meta-linguistic knowledge than their lower-level peers. The hypothesis that indirect CF may be more helpful for learners with higher levels of meta-linguistic awareness

---

11 The fact that pupils in the self-correction group were able to improve their non-grammatical but not their grammatical accuracy during revision (S2) also suggests that the grammatical errors are the more complex ones.
(e.g. Ferris, 2004; Hyland and Hyland, 2006), then led us to predict that the higher-level pupils might be better able to use indirect corrections. Although we did find that the higher-level pupils outperformed the lower-level learners on the different linguistic measures used in the study, we never found a significant interaction between the effectiveness of the CF treatments and learners’ educational level.

4.7.5 Limitations and directions for further research

Although findings of the present study constitute clear evidence in favor of written comprehensive CF, we should keep in mind that it is limited in several ways. To begin with, our participants only received feedback once. Even though the pupils in the indirect group received instruction on how to use and interpret the error codes, they might not have had enough opportunity to get used to this type of correction. It is therefore possible that the effect of indirect CF would have been greater if we had applied a more longitudinal design with more than one CF occasion. In the same line of reasoning, we would not argue that practicing writing does not have any value for accuracy development; it could well be that practicing writing does lead to improved accuracy when learners are allowed more than just one additional practice opportunity. However, our results do not indicate that error correction is a waste of time in the sense that practicing writing is more beneficial (Truscott, 1996; 2004). On the contrary, even a single CF treatment proved to have long-lasting positive effects. These effects could only be expected to be greater if learners are offered CF on additional occasions.

Secondly, the reason that we did not find a significant interaction between pupils’ educational level and CF efficacy, might be that the difference between the levels included in this study was not big enough. It could be that a comparison of more dissimilar groups would have led us to other conclusions concerning the influence of educational background on feedback benefits. It also needs to be noted that the present study was performed within a particular context, which could be described as a relatively naturalistic SLA environment; pupils started learning Dutch at an early age, and the L2 was the means rather than the goal of instruction. As a result, our learners’ level of meta-linguistic awareness might have been too low to really and fully benefit from indirect CF, even for the higher-level pupils.

Thirdly, even though the present investigation was performed in intact classes, it could not be considered a real classroom study. The administered tasks – although representative of the written assignments used within the educational setting under investigation – were not part of the curriculum, and the feedback and task instructions were provided by the researcher instead of by the teacher. Consequently, we cannot be sure that
our findings would hold in a real-world class situation. On the other hand, we would not have been able to unambiguously establish the effectiveness of comprehensive CF in a less controlled setting.

Moreover, one could question our operationalization of grammatical accuracy. We included article errors, inflectional errors, word order errors, omissions, additions, and pronominal errors in our broad measure of grammatical accuracy. The syntactic errors concerning word order, and additions or omissions of constituents are undoubtedly problems that involve integral parts of a complex system, and therefore certainly belong to the type of grammar errors Truscott claimed (2007) could never benefit from CF. In comparison, errors in the use of articles, pronominals, and inflectional morphology might be considered less complex problems in relatively discrete items. However, analyses on a narrow category of syntactic errors only (i.e. word order errors, additions of non-necessary elements, and omissions of necessary elements) rendered the exact same results as the analyses on the broader measure of grammatical accuracy reported in this article. From this we conclude that comprehensive CF is effective in decreasing the number of even the most complex types of grammatical errors. Further research is warranted, however, to identify the exact effect of comprehensive CF on separate types of grammar problems, or even different functional uses of a single grammatical feature.

Finally, as Bruton (2009a) argued, comparing the error rates of two texts might not be the best way to investigate if and how learners have benefited from CF. In doing so one “assumes that the experimental group students could potentially draw on language knowledge in the second writing task gained from the feedback on the first writing task” (Bruton, 2009a, p. 137). It is uncertain, however, if learners can indeed use the knowledge they have gained from corrections on a particular piece of writing in a new text, on a new topic, in a new genre, and so on. Although our tasks were designed to be as similar as possible, it is still quite feasible that participants received corrections on features that they were unable to reuse when writing a new text. More detailed, qualitative analyses of learners’ accuracy performance over time might give a better picture of the accuracy gains brought about by CF. Research that explores this potential added value of an in-depth accuracy analysis would be worthy of further pursuit (cf. Chapter 5).

4.7.6 Conclusion

Despite the limitations listed above, our results clearly showed comprehensive CF to be effective in promoting both grammatical and non-grammatical accuracy during revision as well as in new pieces of writing, irrespective of learners’ educational level. Moreover, our
results did not constitute evidence of learners avoiding complex structures due to CF, nor did we find writing practice to be more beneficial than CF. Hence, the present study does not support Truscott’s (1996; 1999; 2004; 2007; 2009) claim that written CF is ineffective or even harmful. We conclude that comprehensive CF is a useful instrument that L2 teachers can use to help learners improve their accuracy in writing.
Chapter 5

An in-depth analysis of the effects of direct and indirect corrective feedback in L2 writing¹

Qualitative study

5.1 Abstract

The value of corrective feedback (CF) for second language (L2) development has been debated ever since Truscott’s (1996) article in which he claimed error correction to be ineffective and potentially harmful. Recent studies have shown, however, that written CF does promote L2 learners’ accuracy development. Nevertheless, little is known about the ways in which individual learners engage with and benefit from the feedback they receive, because most studies assessed accuracy gains quantitatively in terms of global measures (e.g. error rates). The present multiple case-study set out to fill this gap by analyzing the accuracy performance of four L2 learners of Dutch in detail. In doing so, we explored the amenability of different error types (e.g. lexical errors, morphosyntactic errors) to direct and indirect CF. Findings proved that in-depth accuracy analysis is a valuable addition to global measurement of accuracy development, because it can reveal details on CF effectiveness that will not be unveiled by global analyses. Our study showed, for example, that the efficacy of CF may be mediated by factors such as learners' level of successful CF uptake, or the nature of the targeted error.

¹ A slightly adapted version of this chapter has been submitted as: Van Beuningen, C. G., De Jong, N. H., & Kuiken, F. (under review). Corrective feedback in L2 writing: An in-depth analysis of the effects of direct and indirect corrective feedback in second language writing.
5.2 Introduction

Although common practice in L2 instruction, the value of CF for accuracy development has been a fiercely debated topic (see particularly Ferris, 1999; 2004; Truscott, 1996; 1999; 2007; Truscott & Hsu, 2008). Truscott (1996) even summoned the abandonment of CF from L2 classrooms. He argued that, based on insights from SLA theory, CF would be more likely to harm than to promote L2 learners’ accuracy development (see Chapters 2 and 4 for a full review of Truscott’s objections against CF). He furthermore stated that the available research base was unable to refute his claim that CF is necessarily ineffective and potentially harmful.

Truscott’s (1996) case against CF gave rise to an ever growing number of studies investigating the effectiveness of error correction. Early CF work mainly focused on the role of CF in the revision process, and demonstrated that correction enables learners to improve the accuracy of a particular piece of writing (e.g. Ashwell, 2000; Fathman & Whalley, 1990; Ferris, 1997; Ferris & Roberts, 2001; Sachs & Polio, 2007). Whereas these studies proved that CF is a useful editing tool, they do not validate any conclusions on the role of CF in the process of L2 learning. Investigating the value of CF for L2 acquisition, would necessarily involve “a comparison between two independently written works” (Truscott & Hsu, 2008, p. 293). Revision studies, however, only compare two versions of the same text.

Earlier studies that did opt to investigate CF’s effectiveness in yielding a learning effect (by looking into the effect of CF on newly written texts), failed to provide conclusive evidence on the value of CF for L2 acquisition because of methodological short-comings² (e.g. Chandler, 2003; Kepner, 1991; Polio, Fleck, & Leder, 1998; Semke, 1984; Sheppard, 1992). A number of recent, tightly controlled investigations, however, has proven that CF does have the potential to foster L2 learning, by showing that CF enables learners to improve the accuracy of new pieces of writing (see Chapters 3 and 4; see also Bitchener, 2008; Bitchener & Knoch, 2008; Bitchener & Knoch, 2009; Bitchener & Knoch, 2010a; Bitchener & Knoch, 2010b; Ellis, Sheen, Murakami, & Takashima, 2008; Sheen, 2007).

Although these recent studies contributed greatly to the error correction debate by tackling the question if CF is an effective means of improving L2 learners’ written accuracy over time, they do not provide insights into how and when learners benefit from error correction, because they have all investigated the effectiveness of written CF by comparing group performances over time, rather than by in-depth comparison of individual learners’

² See Chapter 3 or, for example, Bitchener (2008) and Guenette (2007) for a review of these design related issues.
sequential accuracy performances. As Storch and Wigglesworth (2010) stated “research that analyzes actual instances of learners engaging with feedback and revising their texts […] is needed to understand how […] learners respond to (different forms of) CF” (p. 304).

Moreover, most of the written CF research – except for the studies investigating the effect of CF on one specific error type (Bitchener, 2008; Bitchener & Knoch, 2008; Bitchener & Knoch, 2009; Bitchener & Knoch, 2010a; Bitchener & Knoch, 2010b; Ellis et al., 2008; Sheen, 2007) – used global accuracy measures to assess CF efficacy (i.e. number of errors per 100 words, number of error-free T-units, etc.). Bruton (2009a) suggested, however, that comparing two texts on such global accuracy measures might not be the best way to investigate if learners benefit from error correction, and argued that more detailed, qualitative analysis of learners’ accuracy performance over time might give a more complete and accurate picture of the accuracy gains brought about by CF. Bruton based this claim on the outcomes of his detailed re-analysis of a portion of Truscott and Hsu’s (2008) data. Comparing the sequential error rates of this study’s CF group (i.e. the experimental group) to those of its control group, led Truscott and Hsu to conclude that the accuracy gained by their CF group during revision did not lead to improved accuracy in a new text. Hence Truscott and Hsu stated that “the successful error reduction during revision [was] not a predictor […] of learning” (p. 299). However, when subjecting a sample of Truscott and Hsu’s data to an in-depth inspection, Bruton did find evidence of a carry-over effect of previous corrections on subsequent writing. He explained that the global error rates used by Truscott and Hsu failed to reveal these observed accuracy gains, because measuring accuracy development by means of global error rate comparison is based on false premises. First of all, it presupposes that a learner will (be able to) use the features that were corrected in his initial text when writing a new one. Secondly, Bruton argued that comparing an initial text and a subsequent piece of writing on global accuracy measures “does not reflect the assumed relationship between the errors in the two texts: it is not that the student writer made some grammatical errors, which were […] corrected in a revision, only to reappear a week later” (p. 139). In reality the sample from Truscott and Hsu’s data showed that a lot of the errors in the subsequent text did not bear any relation to the errors corrected in the first piece of writing.

Another drawback of global accuracy measurement identified by Bruton (2007; 2010) is that, in adopting all-or-nothing criteria (i.e. accurate/inaccurate), it will inevitably fail to quantify partial learning. It is conceivable, however, that a learner will show retention for only part of a particular feedback instance (as will also become clear from our findings).

A final issue into which qualitative accuracy analysis could provide valuable insights,
is the potentially differential effects of various CF forms on distinctive types of errors. It has been argued, that no single form of CF should be expected to be effective in addressing all types of linguistic errors (e.g. Ferris, 1999; Truscott, 1996), because morphological, syntactic, and lexical errors represent gaps within different domains of linguistic knowledge (e.g. Schwartz, 1993). In-depth analyses of learners’ writing would enable exploration of possible interactions between types of errors and CF methodologies.

Two CF methodologies that have received a lot of attention, are direct CF (i.e. both errors and target forms are indicated by the teacher) and indirect CF (i.e. errors are indentified without provision of target forms)\(^3\). While some researchers have suggested that learners will benefit more from indirect CF because they have to engage in a more profound form of language processing when they are self-editing their writing (e.g. Ferris, 1995; Lalande, 1982), we (cf. Chapter 4) and others have claimed that the indirect approach might fail because indirect CF provides learners with insufficient information to resolve complex linguistic problems such as syntactic errors (e.g. Bitchener & Knoch, 2010b; Chandler, 2003).

Only two studies that we are aware of, investigated the interaction between error type and the effectiveness of direct and indirect CF methodologies. An explorative study by Ferris (2006) showed that 86 ESL students only realized a significant reduction over the semester in errors concerning verb form, and that the majority of this kind of errors was corrected indirectly. These observations led Ferris to suggest that indirect CF was more beneficial to accuracy development than direct CF. However, this study was not initially designed, and therefore unable to directly compare the two CF methodologies. In fact, Ferris set out to explore the value of indirect CF only, but found that the teachers in her study addressed different types of errors with different forms of CF.

The study reported on in Chapter 4, investigated the effectiveness of direct and indirect comprehensive CF in a tightly controlled classroom-based study (N = 268) incorporating pre-test, treatment, and (delayed) post-test sessions. This study contrasted two experimental treatments (i.e. direct CF and indirect CF) and two control treatments (i.e. self-correction and writing practice). Results not only showed comprehensive CF to be effective in promoting accuracy over time, they also revealed an interaction between error type and CF methodology: whereas direct CF proved to be most effective in remedying grammatical errors (i.e. morphosyntactic errors), learners’ improvement on non-grammatical features (e.g. spelling errors, punctuation errors) was retained the longest when indirect corrections were provided.

---

\(^3\) See Chapter 4 for a full review of the studies contrasting direct and indirect CF types.
The findings presented in Chapter 4 contribute to the error correction debate (e.g. Truscott, 1996; Ferris, 1999) by refuting Truscott's (e.g. 2007) claim that grammatical errors are insusceptible to CF, and by showing that direct correction might be most beneficial to learners’ development of grammatical accuracy. A limitation, however, lies in the broadness of the contrasted error categories (i.e. grammatical vs. non-grammatical errors). It might well be that CF types interact differently with separate error types within these broad domains (see also Bitchener, 2008). In-depth analysis of learners’ accuracy performance over time has the potential to reveal such fine-grained interdependencies between the effectiveness of CF forms and error types.

5.3 The present study

5.3.1 Aims
The present multiple case-study adopted detailed, sequential error analyses to explore the effects of written CF on L2 learners’ accuracy development. In doing so, we took the above mentioned reservations towards global accuracy measurement at heart, and acted upon Storch and Wigglesworth’s (2010) call for more qualitative CF studies. With the in-depth approach we aimed at (1) furthering our understanding on how and when L2 learners benefit from written CF, and at (2) providing insights into the (potentially) differential effects of direct and indirect CF on different types of errors. The cases under investigation were taken from the larger quantitative project presented in Chapter 4, in which 268 participants took part.

5.3.2 Setting and participants
From the 268 participants in the larger project, four were selected for closer examination in the present multiple case-study. Participants were pupils in their second year of Dutch pre-vocational secondary education (N = 134) or higher general secondary education (N = 134). They attended schools with multilingual student populations. Over 80% of those schools’ pupils came from non-Dutch language backgrounds; most pupils were born in the Netherlands, but many of them only started learning Dutch in school (i.e. at age four).
All four case-study participants received either direct or indirect CF during the treatment session. For each feedback type, we selected one pupil whose error rate decreased after he or she had received CF, and one pupil whose error rate increased between pre-test and post-tests. By selecting pupils who seemingly reacted very differently to the CF they received, we hoped to gain a better understanding of why, how, and when learners (fail to) benefit from correction.

The case-study participants furthermore satisfied the following selection criteria: (i) Dutch was not their L1; (ii) they participated in all four experimental sessions; (iii) they showed an average (within one SD of sample mean) performance on the pre-test measure of written accuracy (i.e. number of linguistic errors per 10 words); (iv) their pre-test writing contained errors within a range of linguistic categories.

### 5.3.3 Treatments and procedure

Participants attended four experimental sessions; a pre-test session (S1), a treatment session (S2), a post-test session (S3), and a delayed post-test session (S4). During the first session (S1) pupils were given 20 minutes to complete the first writing task. They were also presented with a receptive vocabulary test (cf. Appendix E), and a questionnaire concerning their language background (cf. Appendix D). Scores on the vocabulary test provided an indication of pupils’ overall language proficiency (e.g. Zareva, Schwanenflugel, & Nikolova, 2005). During the treatment session (S2), which took place one week after S1, pupils received the corrected versions of their initial text. The first experimental group was provided with direct comprehensive CF, which consisted of identification of all existing linguistic errors and provision of the corresponding target forms. The second experimental group received comprehensive feedback in the form of indirect corrections, consisting of indications of errors and error codes corresponding to the relevant error categories. It was left to the student to derive the target forms. Pupils in both groups were instructed to copy their initial text while revising all errors corrected by the researcher, and were allocated 20 minutes to finish the task. The first post-test (S3) was administered one week after the treatment session (S2), and the delayed post-test (S4) took place one month after S2.

---

4 Since the quantitative study presented in Chapter 4 already showed that receiving CF was more beneficial to accuracy development than doing without, this case-study’s focus was only on exploring the effects of CF more in-depth. Therefore, the present study does not include pupils who received one of the control treatments (i.e. self-correction or writing practice).
During both post-test sessions, pupils were given 20 minutes to produce a newly written text\(^5\).

### 5.3.4 Writing tasks

The tasks used throughout the experiment were three writing assignments on the metamorphosis of different insects, that is butterflies (S1, S2), ladybugs (S3), and honey bees (S4). All tasks within the series had a comparable form and set-up; they invited pupils to write an e-mail to a classmate explaining the metamorphosis of the particular insect, based on a series of images depicting the metamorphosis process. (Cf. Appendix A.)

### 5.3.5 Data analysis

As did the studies in Chapters 3 and 4 and other investigations into the effectiveness of written CF (e.g. Chandler, 2003; Truscott & Hsu, 2008), we calculated and compared error ratios (i.e. \((\text{number of form-related errors/total number of words}) \times 10\)) to be able to measure learners’ accuracy development. Because of the drawbacks associated with this type of global assessment of accuracy improvement (e.g. Bruton, 2007; 2009a; 2010; Storch & Wigglesworth, 2010), however, we also used more detailed measures to establish pupils’ levels of successful CF uptake and retention.

We compared the texts pupils wrote during the pre-test session (S1) with the revised versions produced in the treatment session (S2), along with the provided corrections, to trace for evidence of successful feedback uptake. Utterances in the initial texts on which CF was provided, were identified in the revised versions as either (a) corrected/reformulated correctly, (b) uncorrected, (c) revised/reformulated incorrectly, or (d) deleted. Since successful uptake can be defined as a learner’s correct response to a CF instance, corrections and correct reformulations were taken to show that CF instances were taken up successfully.

Texts written throughout the post-test (S3) and delayed post-test sessions (S4) were analyzed for proof of accuracy development – or learning – brought about by CF. Accuracy development was operationalized as the target-like usage of a corrected feature in a new piece of writing (i.e. texts written during S3 or S4).

Since CF could not be expected to have any effect on features it did not target (Bruton, 2009a), we also considered it important to identify the errors in learners’ revisions

---

\(^5\) See Appendix B for the instructions the different groups received during the treatment session (S2), and Appendix C for a handout on the error coding system pupils in the indirect CF group were provided with.
(S2), post-test texts (S3), and delayed post-test writing (S4) that did not bear any relation to the pre-test (S1) problems they received CF on. We classified such errors as ‘new errors’. It needs to be noted that a relation between two errors does not imply a one-to-one correspondence; when, for example, a learner received feedback on an incorrect agreement marker on a certain verb in his initial text, and he committed the same agreement error on a different verb in a subsequent piece of writing, we did not classify this latter error as a new problem. In this case, we considered the provided CF to be readily transferable to other but similar structures and utterances.

To be able to explore the amenability of different types of errors to CF, the errors in pupils’ writing were classified into four broad linguistic domains, that is morphosyntax, lexicon, orthography, and pragmatics. Within the different linguistic categories errors were subsequently subdivided into 11 separate error types (cf. Table 5.1).

<table>
<thead>
<tr>
<th>Linguistic domain</th>
<th>Error type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Morphosyntax</td>
<td>Word order error</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Omission of a necessary element</td>
<td>Omission of constituent (e.g. subject, object, verb)</td>
</tr>
<tr>
<td></td>
<td>Addition of a non-necessary element</td>
<td>Addition of superfluous constituent</td>
</tr>
<tr>
<td></td>
<td>Determiner error</td>
<td>Omission/superfluous determiner, incorrect determiner</td>
</tr>
<tr>
<td></td>
<td>Referential error</td>
<td>Lack of number/gender agreement pronoun - referent</td>
</tr>
<tr>
<td></td>
<td>Inflectional error</td>
<td>Incorrect verbal, nominal, adjectival inflection</td>
</tr>
<tr>
<td>Lexicon</td>
<td>Word choice error</td>
<td>-</td>
</tr>
<tr>
<td>Orthography</td>
<td>Capitalization error</td>
<td>Omission/superfluous use of upper-case character</td>
</tr>
<tr>
<td></td>
<td>Punctuation error</td>
<td>Omission/incorrect use of punctuation mark</td>
</tr>
<tr>
<td></td>
<td>Spelling error</td>
<td>-</td>
</tr>
<tr>
<td>Pragmatics</td>
<td>Contextual error</td>
<td>Error concerning politeness, pronoun without formerly introduced referent</td>
</tr>
</tbody>
</table>
5.4 Findings

In this section, we present the observations about the accuracy performance of our four case-study participants: Emre, Nathalie, Mehmet, and Dinesh. For each of the participants, we first provide some background information, and then consecutively analyze their sequential error rates, level of CF uptake, and level of CF retention. We conclude each case-study by summarizing the most important observations.

5.4.1 The case of Emre: indirect CF and decreasing error rates

Background
Emre is a 14 year-old male, who was born in the Netherlands. Both of his parents were born in Turkey, and speak Turkish with Emre. He attended the second year of higher general secondary education at the time of data collection. Emre’s overall language proficiency in Dutch – which was measured by means of a receptive vocabulary test – was below average compared to that of pupils at the same educational level; he answered 64 out of 108 items correctly, while the average score for his level of education was 76.5 (SD = 10.10).

Sequential error rate analysis
Table 5.2 shows the number of errors and their distribution over different error categories in the texts Emre wrote during pre-test (S1), treatment (S2), post-test (S3), and delayed post-test (S4) sessions.

Emre’s initial text (i.e. the text written during S1) contained 23 errors, on which he received indirect CF in the form of correction codes. The overall error rate (i.e. (number of form-related errors/total number of words) x 10) for this text was 2.32.

The sequential error rates of Emre’s texts show that he was able to considerably improve his initial text through revision, and indicate that he still benefited from the CF when writing a new text; at S2 Emre was able to bring back the number of errors in his first text from 23 (S1) to 4, rendering a reduction of 1.51 errors per ten words (i.e. from 2.32 at S1 to

---

6 The names of the case-study participants reported on in this chapter are fictitious.

7 It was explained in Chapter 3 that it might not be valid to directly compare pupils’ accuracy performance on different writing tasks, because tasks might differ in difficulty. We believe, however, that the qualitative perspective of the study presented in this chapter allows for such a sequential accuracy analysis; it was adopted to examine learners’ accuracy development in relation to the sequential accuracy performances of the other case-study participants.
0.81 at S2). When comparing the number of errors in Emre’s pre-test text \((n = 23)\) to his performance on the post-test \((n = 11)\) and delayed post-test \((n = 9)\), we see that CF also constituted improved accuracy over time. Emre realized an error rate reduction of 1.36 errors between the pre-test and the two post-tests (from 2.32 at S1 to 0.96 at S3 and S4).

Table 5.2 furthermore shows that 64% of the errors Emre committed at S3 and 44% of the errors he made at S4 were totally new; they did not hold any relation to the feedback provided at S2.

<table>
<thead>
<tr>
<th>Error type</th>
<th>No. of occurrences S1</th>
<th>No. of occurrences S2</th>
<th>No. of occurrences S3</th>
<th>No. of occurrences S4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Word order errors</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Omissions of necessary elements</td>
<td>0</td>
<td>0</td>
<td>1(^N)</td>
<td>1(^N)</td>
</tr>
<tr>
<td>Additions of non-necessary elements</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Determiner errors</td>
<td>5</td>
<td>1</td>
<td>2 (1(^N))</td>
<td>0</td>
</tr>
<tr>
<td>Referential errors</td>
<td>0</td>
<td>0</td>
<td>1(^N)</td>
<td>1(^N)</td>
</tr>
<tr>
<td>Inflectional errors</td>
<td>3</td>
<td>0</td>
<td>1(^N)</td>
<td>2 (1(^N))</td>
</tr>
<tr>
<td>Word choice errors</td>
<td>1</td>
<td>1</td>
<td>1(^N)</td>
<td>1(^N)</td>
</tr>
<tr>
<td>Capitalization errors</td>
<td>9</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Punctuation errors</td>
<td>5</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Spelling errors</td>
<td>0</td>
<td>0</td>
<td>2(^N)</td>
<td>0</td>
</tr>
<tr>
<td>Contextual errors</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

**Total no. of errors (no. of words)**: \(23\ (99)\), \(4\ (103)\), \(11\ (114)\), \(9\ (94)\)

\(^N\)New error: not present at S1 and no CF provided in relation to this error

**Analysis of feedback uptake**

Emre showed a high level of successful feedback uptake; he corrected or correctly reformulated the vast majority of the utterances containing an error (i.e. 19 out of 23, or 83%) during revision (S2). The errors that were left uncorrected are one out of nine capitalization errors, one out of five punctuation errors, one out of five determiner errors, and one word choice error (i.e. inappropriate preposition).

Emre proved to be able to use the indirect CF instances targeting errors in the orthographical domain (i.e. capitalization and punctuation errors), and errors within the domain of morphosyntax (i.e. errors in the use of determiners and nominal inflection). He showed no uptake, however, in relation to the feedback provided on the only lexical error in
his initial text, concerning the use of the incorrect preposition *naar* (to) instead of *in* (in). The same error reappeared in Emre’s revised text (S2), as can be seen in excerpt (1):

(1) S1: veranderen *naar* larven\(^8\)  
change  *to* larvae 
[change into larvae]  

S2: veranderen *naar* larven \(^9\)  
change  *to* larvae 
[change into larvae]

This lack of CF uptake might be related to an interaction between the idiosyncratic nature of this lexical error and the indirect feedback Emre received. Ferris (1999; 2002) hypothesized that chances are slim for learners to succeed in self-correcting errors based on indirect CF, when a clear rule behind the solution is lacking. She reasoned that indirect CF provides a learner with too little information to enable him to deduce the target form. Ferris (1999) therefore recommended teachers to give direct CF on such non-rule-based errors, hoping to, “if nothing else, provide input for the acquisition of these idiomatic forms” (p. 6)\(^10\).

Based on these suggestions by Ferris (1999; 2002), we propose that, whereas Emre was able to self-edit his rule-based errors concerning capitalization, punctuation, determiners, and inflection, the indirect CF he received failed to provide him with sufficient information to solve his non-rule-governed lexical error in (1).

**Analysis of feedback retention**
Emre’s sequential error rates already showed that he was able to improve the accuracy of his writing over time. However, a closer look at two corrected constructions that reappear in

---

\(^8\) The asterisk in front of an utterance indicates an erroneous construction.  

\(^9\) In the original text the noun *larve* (*larva*) did not agree in number with the verb *veranderen* (*to change*) (cf. excerpt (3)). To enhance their saliency, however, the excerpts in this chapter only present one error at a time. Errors which were irrelevant for the example at hand were corrected.

\(^10\) Ferris (1999; 2002) labeled non-idiomatic or idiosyncratic errors as “untreatable” errors. Errors that occur in a patterned and rule-governed way, on the other hand, were categorized as “treatable”. In our opinion this terminology is unfortunate. It seems to propagate the rather fatalistic idea that it is impossible to remedy learners’ non-rule-based errors, irrespective of the type of feedback provided. However, the way Ferris related the nature of errors to their amenability to direct and indirect CF seems to be useful; whereas rule-governed errors may be treatable by indirect CF, idiosyncratic errors appear to be untreatable by indirect correction.
the texts Emre produced during the post-test (S3) and delayed post-test sessions (S4), lead us to interesting observations that stay hidden behind the numbers.

The first construction from Emre’s writing we will examine in more detail, concerns a re-emerging (incomplete or erroneous) determiner-noun combination. In Dutch, nouns in singular contexts must be preceded by an article. Dutch distinguishes between definite and indefinite determiners, and has a two-way gender system (i.e. neuter gender and common gender) which surfaces, for example, in gender agreement between the noun and the definite article; whereas common gender nouns must be combined with de (the), neuter gender nouns are accompanied by the article het (the)\(^{11}\).

Excerpt (2) shows that Emre omitted the article which should have preceded the noun metamorfose (metamorphosis) at S1. The feedback he received on this error enabled him to correctly revise the utterance at S2. In the text Emre wrote during the first post-test session (S3), he produced the same noun phrase, this time including the article. However, the article het (the) and the noun metamorfose (metamorphosis) do not agree in gender (the article having neuter, the noun common gender), rendering an error in the determiner category. Finally, during the delayed post-test (S4), Emre did combine metamorfose (metamorphosis) with the agreeing determiner de (the).

(2) S1: Het ging over ø-DET metamorfose-COM SG van een vlinder
     \[It was about ø-DET metamorphosis of a butterfly\]
     [It was about the metamorphosis of a butterfly]

S2: Het ging over de metamorfose-COM SG van een vlinder
    \[It was about the metamorphosis of a butterfly\]

S3: Het gaat over *het metamorfose-COM SG van een lieveheersbeestje
    \[It is about *the metamorphosis of a ladybug\]

S4: … vertellen over de metamorfose-COM SG van de wesp
    \[… tell about the metamorphosis of the wasp\]

---

\(^{11}\) Dutch determiners only distinguish between the two genders in definite, singular contexts. In plural contexts de is used in combination with both common and neuter gender nouns. In indefinite contexts all nouns – irrespective of their gender – combine with the article een (a(n)).
A comparable pattern of development that is visible from Emre’s writing, concerns the plural formation of the noun larve (larva). Dutch has two options for pluralizing root nouns; the plural form can be created by attaching either the affix -(e)n (/l(ə)n/) or -s (/s/). The criteria for determining which affix attaches to which noun, are phonological in nature. One of the relevant features is stress\textsuperscript{12}; whereas a word with a stressed final syllable usually combines with –en, a word ending in an unstressed syllable will prefer –s (Van Wijk, 2002).

Rules governing the plural formation of simplex nouns ending in /\textauml/ are less straightforward. While the majority of words in this class correctly combines with both affixes (-n and -s), some exclusively take either -n or -s. As becomes clear from excerpt (3), Emre struggled with the pluralization of such a noun, namely larve (larva). This noun can only be combined with the plural affix -n.

\begin{verbatim}
(3) S1: ... *larve-3rd SG zijn-3rd PL ...
       ... *larva are ...
       [...] larvae are ...

       Ze veranderen-3rd PL in *larve-3rd SG ...
       They change into *larva ...
       [They are changing into larvae]

S2: ... larven-3rd PL zijn-3rd PL ...
       ... larvae are ...
       [...] larvae are ...

       Ze veranderen-3rd PL in larven-3rd PL ...
       They change into larvae ...
       [They are changing into larvae]

S3: Ze groeien uit tot een-DET SG larve-SG
       They grow out to a larva
       [They grow out to be a larva]

S4: Ze groeien-3rd PL uit tot *larves-*3rd PL
       They grow out to * larvae
       [They grow out to be larvae]
\end{verbatim}

\textsuperscript{12} Factors such as sonorancy of the final consonant and the number of consonants in the final coda are also involved in Dutch noun pluralization. A comprehensive description of Dutch plural formation is beyond the scope of this thesis, but see, for example, De Haas & Trommelen (1993) for details.
In his initial text (S1), Emre failed to realize pluralization of the noun *larve (larva)* twice, resulting in two subject-verb agreement errors; he combined the singular form of the noun *larve (larva)* with the plural form of the verbs *zijn (to be)* and *veranderen (to change)*. Just as in example (2), Emre succeeded in producing the target plural form during revision (S2). At S3 he correctly combined the singular noun with the singular form of the verb. In the delayed post-test Emre reintroduced the plural form of the noun *larve (larva)*, albeit by using an incorrect plural affix –s.

The particular patterns in Emre’s use or failure to use the (correct) determiner and plural marking (i.e. nominal inflection) in excerpts (2) and (3), provide a better understanding of how learners might use indirect CF, and offer insights into the amenability of different types of errors to this kind of feedback. How can it be explained that Emre was able to self-edit the errors in (2) and (3) correctly during revision, failed to use the correct form in one of the post-tests, but produced the target form in the other? To answer this question we again resort to Ferris’ (e.g. 1999, 2002) prediction that rule-based and non-rule-based errors differ in their amenability to indirect CF. In the following paragraphs we suggest that the indirect CF Emre received, only partially remedied the errors presented in excerpts (2) and (3).

Although excerpt (2) shows that Emre committed a determiner error in the same noun phrase before and after receiving CF, the S1 error is different in nature from the error in the post-test text (S3). Whereas Emre used the wrong determiner at S3, he omitted the determiner altogether in his S1 writing. This determiner omission at S1 did not stand on itself. All determiner errors in his initial text were cases of article omission; Emre omitted five out of seven obligatory determiners during the pre-test13. Conversely, in his post-test writing, Emre realized all obligatory determiners (i.e. 8 at S3, and 10 at S4). The only remaining determiner errors concerned the use of an incorrect article.

We propose that the indirect CF Emre received, might have raised his awareness of the rule which dictates Dutch singular count nouns to be accompanied by a determiner. It could be assumed, that this awareness then led Emre to add an article during revision, and to not omit any articles in his (delayed) post-test writing. However, although Emre chose the correct article *de (the)* when revising, he could not be sure whether his hypothesized choice for this determiner was indeed accurate (see also Chandler, 2003); there is no straightforward rule in Dutch prescribing whether a noun has neuter or non-neuter gender. This uncertainty might have prevented Emre from internalizing the correct form. This

---

13 The other four determiner omissions concerned different article-noun combinations than the one presented in excerpt (2).
clarifies how it was possible for Emre to incorrectly combine metamorfose (*metamorphosis*) with *het* (*the*) in the post-test, and again correctly with *de* (*the*) in the delayed post-test. Whereas the error concerning the rule-governed use of Dutch determiners proved to be susceptible to indirect CF, his error pertaining to the idiosyncratic determiner-noun combination was not. Until Emre is provided with direct input on which article accurately combines with *metamorfose* (*metamorphosis*), he has to choose between the two available options (i.e. *de* and *het*) every time he produces the particular noun.

Just as the use of determiners, the plural formation of nouns is only partially governed by a clear-cut rule in Dutch, and therefore only partially receptive to indirect CF. In general, simplex nouns with an /s/ ending can take both plural markers *-n* and *-s*. Nevertheless, *larve* (*larva*) is one of the exceptional nouns which only allows combination with *-n*. This makes that, for this noun, pluralization is a more idiosyncratic issue, and explains Emre’s sequential performance in excerpt (3). During S1, he failed to use any plural marking on the noun *larve* (*larva*). We suggest that – when provided with indirect CF – Emre became aware of the requirement for subjects to agree in number with their matching verbs. Due to this awareness, no omissions of nominal plural affixes remained in the two post-tests. Nonetheless, indirect CF did not succeed in fully remedying Emre’s problems with plural morphology. Parallel to excerpt (2), Emre selected the correct plural marker *–n* during revision (S2), but the incorrect one (i.e. *–s*) in the delayed post-test session (S4).

**Summary**

A comparison of Emre’s sequential error rates revealed evidence of feedback uptake and retention; after receiving indirect CF, Emre produced substantially fewer errors in both his revised text, as well as in his newly written texts. It also showed that about half of the errors in Emre’s post-test writing did not bear any relation to the CF he received. Exploring Emre’s accuracy performance at a more detailed level, however, provided additional insights into the amenability of different error types to indirect CF, and revealed evidence of partial accuracy improvement that was not represented in the overall error rates.

**5.4.2 The case of Nathalie: direct CF and decreasing error rates**

**Background**

Nathalie is a 14 year-old female, born in the Netherlands, whose family is from Surinamese origin. Her L1 is Sranan Tongo, a creole language originally spoken by the Creole
population of Suriname. At the time of data collection, Nathalie attended the second year of higher general secondary education. With a score of 91 out of 108 on the vocabulary test, Nathalie’s proficiency in Dutch was above average (i.e. 76.5 items correct, SD = 10.10) in comparison to pupils with the same educational background.

**Sequential error rate analysis**

Table 5.3 presents the number and distribution of errors in Nathalie’s texts. She produced a total of 21 errors in the text she wrote during S1, rendering an overall error rate of 1.71 errors per 10 words. Natalie received direct CF on these errors. The feedback enabled her to reduce the number of errors in her initial text from 21 to 3 during revision (S2). As a result, the overall error rate dropped by 1.47 errors (i.e. from 1.71 at S1 to 0.24 at S2).

A comparison of the number of errors Nathalie committed in her initial text ($n = 21$) to her performance on the post-test (S3) ($n = 7$) and delayed post-test (S4) ($n = 5$), suggests that the direct CF Nathalie received also positively influenced the accuracy of newly written texts; she realized a reduction in error rate of 1.07 errors per 10 words between the pre-test and the first post-test (from 1.71 at S1 to 0.64 at S3), and of 1.21 errors per 10 words between S1 and S4 (from 1.71 at S1 to 0.50 at S4).

Moreover, as illustrated by Table 5.3, 43% of the errors at S3, and 40% of the S4 errors, were new in the sense that they did not bear any relation to the CF Nathalie received.
Table 5.3. Number and types of errors – Nathalie

<table>
<thead>
<tr>
<th>Error type</th>
<th>No. of occurrences S1</th>
<th>No. of occurrences S2</th>
<th>No. of occurrences S3</th>
<th>No. of occurrences S4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Word order errors</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Omissions of necessary elements</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1^N</td>
</tr>
<tr>
<td>Additions of non-necessary</td>
<td>0</td>
<td>0</td>
<td>1^N</td>
<td>0</td>
</tr>
<tr>
<td>elements</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Determiner errors</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Referential errors</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Inflectional errors</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Word choice errors</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Capitalization errors</td>
<td>2</td>
<td>0</td>
<td>1^N</td>
<td>1</td>
</tr>
<tr>
<td>Punctuation errors</td>
<td>12</td>
<td>3</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Spelling errors</td>
<td>0</td>
<td>0</td>
<td>2^N</td>
<td>1^N</td>
</tr>
<tr>
<td>Contextual errors</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Total no. of errors (no. of words)</td>
<td>21 (123)</td>
<td>3 (129)</td>
<td>7 (110)</td>
<td>5 (101)</td>
</tr>
</tbody>
</table>

^N New error: not present at S1 and no CF provided in relation to this error

Analysis of feedback uptake

As did Emre, Nathalie showed a high level of successful CF uptake; she incorporated 18 out of 21 corrections (i.e. 86%) in the revised version of her initial text. In contrast to Emre, however, Nathalie received direct feedback on the errors she committed. Whereas Emre was presented with the challenging task of correcting his own errors using the error codes in his text, Nathalie only had to copy the corrections provided by the researcher. The fact that she was able to perform this task adequately in 18 out of 21 instances is therefore hardly surprising. One might even question why not all corrections led to uptake, leaving three punctuation errors after revision (two of which were left uncorrected while the other one was revised incorrectly). We suggest that Nathalie might have failed to incorporate two of the provided CF instances due to the relatively low saliency of corrections concerning punctuation, which possibly led her to simply overlook them.

Analysis of feedback retention

Global error rate analysis revealed that Nathalie’s (delayed) post-test writing was more accurate than her initial text. When exploring her accuracy gains across different error categories (cf. Table 5.3), a first observation is that Nathalie showed a lot of improvement in the use of punctuation marks. Her initial text contained 12 punctuation errors: Nathalie used one incorrect punctuation mark, and punctuation marks were missing altogether in 11 out of
28 obligatory occasions. In the text written during the first post-test she committed only three punctuation errors, two of which involved the use of incorrect punctuation marks. At S3 punctuation was failing in only 1 out of 21 obligatory occasions. Finally, during the delayed post-test Nathalie failed to realize punctuation in just two out of 22 obligatory occasions.

Table 5.3 furthermore shows that Nathalie gained accuracy in the grammatical domain; whereas her pre-test text contained seven morphosyntactic errors, only one such error remained in her post-test and delayed post-test writing. The most interesting observation pertains to one of her S1 inflectional errors. A close look at two reappearing constructions in Emre's writing revealed that non-rule-based errors might be insusceptible to indirect feedback forms. Ferris' (1999) suggestion that such idiosyncratic errors might benefit more from direct correction seems to be supported by the finding that Nathalie did succeed in acquiring an irregular plural affix based on the direct CF she received.

As explained earlier, Dutch nouns are usually pluralized by adding affixes -s and/or -(e)n. Some particular classes of nouns, however, have to be combined with other markers to realize plural formation. One of these irregular affixes is -eren. The noun blad (leaf) belongs to the class of exceptional nouns that combine with this idiosyncratic affix.

Excerpt (4) shows how Nathalie failed to use accurate plural marking on the noun blad (leaf) in the pre-test (S1). The direct feedback she received enabled her to revise the error correctly (S2). She also succeeded in using the correct form twice in a new text (S3).

(4)  

S1: … vol met *blader-PL …  
... full of *leaf ...  
[... full of leafs]  

S2: … vol met bladeren-PL ...  
... full of leafs ...  
[... full of leafs]  

S3: ... vol met bladeren-PL ...  
... full of leafs ...  
[... full of leafs]  

... zich volgegeten met bladeren-PL ...  
... itself stuffed with leafs ...  
[... stuffed itself with leafs]
Summary
Comparing the error rates of Nathalie’s texts supplied evidence of both uptake and retention of direct CF; the revised version of her initial text, as well as Natalie’s post-test writing, contained considerably fewer errors than the piece of writing she produced during the pre-test session. Moreover, more than 40% of the errors Nathalie committed during the post-tests proved to be unrelated to the corrections she was provided with. An in-depth analysis of Natalie’s sequential accuracy performance furthermore highlighted how direct CF succeeded in promoting the accurate use of an idiosyncratic plural affix. Since Emre’s case showed that a similar error was not amenable to indirect CF, we reiterated Ferris’ (1999) suggestion that idiosyncratic errors might best be corrected directly.

5.4.3 The case of Mehmet: indirect CF and increasing error rates

Background
Mehmet is a 14 year-old male, with a Turkish language background; although he was born in the Netherlands, Turkish is the language of communication at home. Mehmet was attending the second year of secondary pre-vocational education when data collection took place. Mehmet answered 55 out of the 108 vocabulary test items correctly, which means that his score, and thus his Dutch language proficiency, are well below average (i.e. 70.8, SD = 10.63) within his educational level.

Sequential error rate analysis
Table 5.4 shows the number and types of errors in Mehmet’s texts. Mehmet committed 23 errors in his initial writing (S1); the overall error rate of this text was 2.25 errors per 10 words.

Contrary to what we saw in Emre’s and Nathalie’s work, Mehmet’s revised text contained more errors \((n = 26)\) than his initial version \((n = 23)\); there was an increase from 2.25 errors per 10 words at S1 to an error rate of 2.47 at S2. Comparison of Mehmet’s pre-test and post-test performances also reveals that Mehmet committed more errors after receiving indirect CF than before; at S3 the error rate of his text was 2.56, and his delayed post-test writing (S4) held 2.53 errors per 10 words.

When we distinguish between errors that already appeared at S1 and errors that did not, we see that 58% of the errors in the revised text (S2), 52% of the errors in the first post-test (S3), and 50% of the errors in the delayed post-test (S4) are totally new; they hold no
relation to the errors committed at S1, and Mehmet did not receive any feedback on these errors.

Table 5.4. Number and types of errors – Mehmet

<table>
<thead>
<tr>
<th>Error type</th>
<th>No. of occurrences S1</th>
<th>No. of occurrences S2</th>
<th>No. of occurrences S3</th>
<th>No. of occurrences S4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Word order errors</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Omissions of necessary elements</td>
<td>1</td>
<td>1^N</td>
<td>2^N</td>
<td>3^N</td>
</tr>
<tr>
<td>Additions of non-necessary elements</td>
<td>1</td>
<td>1^N</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Determiner errors</td>
<td>0</td>
<td>1^N</td>
<td>1^N</td>
<td>1^N</td>
</tr>
<tr>
<td>Referential errors</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1^N</td>
</tr>
<tr>
<td>Inflectional errors</td>
<td>4</td>
<td>3 (2^N)</td>
<td>3 (2^N)</td>
<td>4 (1^N)</td>
</tr>
<tr>
<td>Word choice errors</td>
<td>0</td>
<td>0</td>
<td>2^N</td>
<td>0</td>
</tr>
<tr>
<td>Capitalization errors</td>
<td>5</td>
<td>5 (2^N)</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Punctuation errors</td>
<td>7</td>
<td>10 (5^N)</td>
<td>8</td>
<td>5</td>
</tr>
<tr>
<td>Spelling errors</td>
<td>5</td>
<td>5 (3^N)</td>
<td>4^N</td>
<td>4^N</td>
</tr>
<tr>
<td>Contextual errors</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><strong>Total no. of errors (no. of words)</strong></td>
<td><strong>23 (102)</strong></td>
<td><strong>26 (105)</strong></td>
<td><strong>23 (90)</strong></td>
<td><strong>20 (79)</strong></td>
</tr>
</tbody>
</table>

^N New error: not present at S1 and no CF provided in relation to this error

Analysis of feedback uptake

When we consider the increase in error rate between the initial version of Mehmet’s text and his revision (i.e. from 2.25 at S1 to 2.47 at S2), the numbers give the impression that Mehmet was unable to interpret and use any of the indirect CF (i.e. error codes) provided to him. However, when we take a careful look at the errors Mehmet committed during S1 and S2, we see that only 11 out of the total 26 errors in his revised text are actually related to errors that were present in the initial version (i.e. the same, unrevised error reappeared, or the initial error was revised incorrectly), and that the other 15 are entirely new. From the 23 utterances containing an error in his pre-test writing, Mehmet was able to correctly revise or reformulate 12 during revision. This leads us to conclude that, while global error rates seemed to show that Mehmet failed to benefit from any of the indirect CF instances at S2, he in fact showed successful uptake for over half of the corrections.

Still, the amount of uptake is relatively low, especially within the orthographical domain; Mehmet left five errors in punctuation, three capitalization errors, and one spelling error uncorrected. One more spelling error was revised incorrectly. As we suggested earlier, the low saliency of (CF instances targeting) punctuation and capitalization errors might
explain the moderate level of uptake within these domains. The lack of uptake concerning the two remaining spelling errors might be related to their persistency, as will become clear from excerpts (6), (7), and (8) in the following paragraph.

**Analysis of feedback retention**

Comparing the error rates of Mehmet’s texts would lead to the conclusion that Mehmet did not benefit from the indirect CF he received. To the contrary: his (delayed) post-test writing seemed to be even less accurate than the text Mehmet produced during the pre-test. However, the in-depth analysis of Mehmet’s sequential performance leads to two important observations.

To begin with, exploration of the degree to which Mehmet was able to use the indirect CF during revision, brought to light that his uptake of the provided feedback was quite low, especially of CF that targeted orthographical errors; ten out of 17 errors concerning punctuation, capitalization, and spelling were still present in the revised version of Mehmet’s text. As it has been suggested that successful uptake might be a predictor of successful retention (e.g. Ferris, 2004), it can be understood why Mehmet’s (delayed) post-test writing still contained a lot of errors in orthography.

Secondly, Table 5.4 shows that half of the errors Mehmet committed during the two post-tests hold no relation to the errors he made at S1. The fact that Mehmet’s error rate was higher after he received feedback than before his errors were corrected, should therefore not be interpreted as evidence against the usefulness of error correction; CF could not be expected to have any influence on errors it did not target in the first place (Bruton, 2009a).

As explained above, no evidence of learning could be anticipated when either uptake of CF is lacking, or no CF has been provided on a certain erroneous construction. Yet excerpt (5) illustrates that when feedback had been given on a certain issue, and successful uptake in relation to this CF instance had been realized, Mehmet did show indications of feedback retention.

(5) **S1:** ... kunnen ze *zich vervellen*  
... *can they *REFL shed their skin*  
[... they can shed their skin]

**superfluous reflexive**

**pronoun**

**target form**

**S2:** ... kunnen ze *vervellen*  
... *can they shed their skin*  
[... they can shed their skin]
CHAPTER 5

S3: Larven vervellen …

Larvae shed their skin …

[Larvae shed their skin]

Excerpt (5) shows how Mehmet combined the non-reflexive verb vervellen (to shed skin) with the reflexive pronoun zich (himself) in his pre-test writing, rendering a syntactic error (i.e. addition of a superfluous element). Subsequently this error was revised correctly, and Mehmet was able to use the corrected utterance adequately at S3.

One more interesting observation we can distillate from Mehmet’s work, relates to his errors concerning inflection and/or (rule-based) spelling. Excerpt (6) shows how Mehmet’s use of the third person singular pronoun het (it) in combination with the first person singular form (or stem) of the verb worden (to turn into) resulted in a subject-verb agreement error. Although he corrected this error when revising his text, the same error reappeared four times in his (delayed) post-test writing. In this case, uptake did not translate into retention.

(6) S1: Daarna *word-1SG het-3SG een pop
    Then *turn into it a pupa
    [It then turns into a pupa]

S2: Daarna wordt-3SG het-3SG een pop
    Then turns into it a pupa
    [It then turns into a pupa]

S3: Daarna *word-1SG een larve-3SG een pop
    Then *turn into a larva a pupa
    [The larva then turns into a pupa]

S4: … *word-1SG het-3SG een larve
    … *turn into it a larva
    [… turn into a larva]

    … *word-1SG het-3SG een jonge wesp
    … *turn into it a young wasp
    [… turns into a young wasp]

    … dat je-2SG snel beter *word-1SG
    … that you soon well *get
    [… that you get well soon]
Why did uptake not result in learning here? The explanation might be found in the nature of the errors in excerpt (6). We propose that instead of labeling them as errors in subject-verb agreement, they could just as well be interpreted as spelling errors. In Dutch the phoneme /t/ can take three different written forms; the graphemes <t>, <d>, and <dt> are all pronounced the same when appearing at the end of a syllable. This makes that the verb forms for first person singular word (turn into) and second/third person singular wordt (turn(s) into) have indistinguishable pronunciations. The incorrect combination of a third or second person singular subject with word (turn into) could therefore be categorized as a spelling error. The errors in excerpt (6) together with several other errors in Mehmet’s texts (cf. excerpts (7) and (8)) suggest that he has a general problem pertaining to the spelling of words that end in the phoneme /t/.

Excerpt (7) shows how Mehmet misspelled two past participles in his initial writing (S1). Both should have ended on <d> rather than on <t>. During revision (S2) Mehmet did not show successful uptake for the CF he received on these errors; he left the first error uncorrected, and incorrectly revised the second one. He then committed a comparable error at S4.

(7) S1: … worden *genoemt  
… are *called 
[... are being called]

… worden *gelegd 
… are "laid 
[... are being laid]

S2: … worden *genoemt 
… are "called 
[... are being called]

… worden *gelegd 
… are "laid 
[... are being laid]

S4: … je had *gevraagt 
…you had "asked 
[... you asked]
Yet another similar error is presented in excerpt (8). Mehmet committed a spelling error in the diminutive form of the noun *brood (bread)*; he used a <t> instead of the correct grapheme <d>.

(8) S4: *brootje*  
*bread roll-DIM*  
[small bread roll]  

spelling error (erroneous diminutive suffix)

Excerpts (6), (7), and (8) show that Mehmet’s errors in the use of the different Dutch graphemes representing the phoneme /t/ are very persistent. We propose that such extremely persistent and potentially fossilized (e.g. Selinker, 1972) errors might be to deeply entrenched in a learner’s interlanguage system to be susceptible to CF. This would then explain why feedback uptake was not followed by retention in excerpt (6).

Summary
Merely looking at Mehmet’s increasing sequential error rates, would have led to the conclusion that indirect CF harmed rather than improved his written accuracy. A detailed analysis of Mehmet’s performance over time showed, however, that (i) more than half of the errors in Mehmet’s post-test writing did not hold any relation to the errors that were corrected in the text he wrote during the pre-test session, and therefore CF could not be expected to influence those errors; (ii) when CF failed to lead to retention, this was either because no uptake had taken place, or because CF targeted an extremely persistent error; (iii) when CF targeted an error that was less entrenched and the feedback instance was successfully taken up during revision, CF did result in the accurate use of reappearing constructions.

5.4.4 The case of Dinesh: direct CF and increasing error rates

Background
Dinesh is a 13 year-old male, whose parents are from Surinamese origin. Dinesh was born in the Netherlands, but the language spoken at home is Sarnami Hindustani. At the time of data collection, he attended the second year of secondary pre-vocational education. Answering 70 of the 108 vocabulary test items correctly, Dinesh’s level of proficiency in Dutch is close to average (i.e. 70.8, SD 10.63) within his educational level.
Sequential error rate analysis

The number and distribution of the errors Dinesh committed during the different experimental sessions are displayed in Table 5.5. The text Dinesh wrote during the pre-test session (S1) contained 30 linguistic errors, which translates into an overall error rate of 1.96 errors per 10 words. By revising his work based on the direct corrections he received at S2, Dinesh brought back the number of errors in his text from 30 to 9. Consequently, the overall error rate went down by 1.29 errors per ten words (i.e. from 1.96 at S1 to 0.67 at S2).

Contrary to what we saw in Emre’s and Nathalie’s cases, Dinesh’s error rate development seems to indicate that he was unable to translate the improvement realized during revision into long-term accuracy gains. Instead of showing a reduction in the number of errors between pre-test and post-tests, Dinesh’s error rates were higher at S3 (2.32) and S4 (3.30) than at S1 (1.96). A comparison of the types of errors Dinesh committed during the different experimental stages, reveals that this increase is particularly related to the number of orthographical errors Dinesh made in his post-test writing (i.e. capitalization, punctuation, and spelling errors).

Moreover, 78% of the errors in his revised text (S2), 33% of the errors in the first post-test (S3) and 38% of the errors in the delayed post-test (S4) did not hold any relation to the errors Dinesh committed in his initial text (S1) or received feedback on during S2.

Table 5.5. Number and types of errors – Dinesh

<table>
<thead>
<tr>
<th>Error type</th>
<th>No. of occurrences S1</th>
<th>No. of occurrences S2</th>
<th>No. of occurrences S3</th>
<th>No. of occurrences S4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Word order errors</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Omissions of necessary elements</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Additions of non-necessary elements</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Determiner errors</td>
<td>0</td>
<td>0</td>
<td>2(^N)</td>
<td>1(^N)</td>
</tr>
<tr>
<td>Referential errors</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Inflectional errors</td>
<td>3</td>
<td>1(^N)</td>
<td>2(^N)</td>
<td>4 (3(^N))</td>
</tr>
<tr>
<td>Word choice errors</td>
<td>2</td>
<td>0</td>
<td>1(^N)</td>
<td>3(^N)</td>
</tr>
<tr>
<td>Capitalization errors</td>
<td>4</td>
<td>0</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>Punctuation errors</td>
<td>10</td>
<td>5 (3(^N))</td>
<td>14</td>
<td>13</td>
</tr>
<tr>
<td>Spelling errors</td>
<td>3</td>
<td>3(^N)</td>
<td>5(^N)</td>
<td>7(^N)</td>
</tr>
<tr>
<td>Contextual errors</td>
<td>0</td>
<td>0</td>
<td>1(^N)</td>
<td>0</td>
</tr>
<tr>
<td><strong>Total no. of errors (no. of words)</strong></td>
<td>30 (153)</td>
<td>9 (135)</td>
<td>33 (142)</td>
<td>37 (112)</td>
</tr>
</tbody>
</table>

\(^N\) New error: not present at S1 and no CF provided in relation to this error
Analysis of feedback uptake

As did Nathalie, Dinesh received direct CF. When comparing the number of errors in Dinesh’s initial text ($n = 30$) to the number of erroneous utterances in the revised version ($n = 9$) (cf. Table 5.5), it seems as if Dinesh was quite capable of using the feedback he received. Especially when keeping in mind that seven out of the nine errors in his revised text (S2) were not yet present in his pre-test writing (S1), and therefore not targeted by CF. These numbers would lead us to conclude that Dinesh only failed to incorporate 2 out of 30 corrections.

However, a closer look at Dinesh’s revised text brings to light that he actually corrected or correctly reformulated just 17 out of the total 30 errors he committed in his pre-test text (S1). Not only were two punctuation errors left uncorrected, but Dinesh also deleted the utterances containing 11 more errors during revision. These deleted utterances held errors concerning punctuation ($n = 5$), capitalization ($n = 2$), omissions ($n = 2$), word choice ($n = 1$), and nominal inflection ($n = 1$). Unlike bare number comparison, a more detailed analysis shows that the level of successful uptake Dinesh displayed was low, especially when taking into account that he received direct CF.

Since gathering process data was not part of the larger research project, we can only speculate on why Dinesh decided to delete some of the utterances he received feedback on instead of correcting them during revision. Earlier studies that did try to tap into learner-internal cognitive processes (by using think-aloud protocols, retrospective interviews, etc.) have shown that feedback uptake is influenced by affective factors such as learners’ motivation, goals, beliefs, and attitudes towards CF (e.g. Bruton, 2010; Goldstein, 2006; Hyland, 1998; Storch & Wigglesworth, 2010; Swain & Lapkin, 2003). We suggest that Dinesh perceived the goal of the revision activity differently from how it was intended by the researcher. Instead of merely copying his initial text and incorporating the corrections, he might have had a broader interpretation of the task as improving his text as he saw fit (see also Storch and Wigglesworth, 2010), deleting and adding complete utterances.

Analysis of feedback retention

Dinesh’s initial text (S1) contained eight sentence level errors (i.e. errors concerning omissions ($n = 6$), word order ($n = 1$), and referential expressions ($n = 1$)). He showed signs of successful uptake for CF instances targeting six out of those eight errors (i.e. he corrected or correctly reformulated the utterances containing these errors during revision (S2)). The utterance holding the two remaining sentence level errors (i.e. two omissions), was deleted from the revised text. It is difficult to trace retention of the corrections targeting
these sentence level errors because the specific corrected constructions do not reappear in Dinesh’s post-test writing. Nevertheless, it is noteworthy that no sentence level errors remain in the texts Dinesh produced during the two post-tests (S3 and S4). Especially the decrease in omission errors (i.e. from six at S1 to zero at S3 and S4) is remarkable; whereas Dinesh omitted a range of different constituents in his initial text (i.e. auxiliary verb, adverb of time, coordinating conjunction, subject, verbal predicate, locative adverb), his post-test writing contained no incomplete sentences.

Another observation pertains to the relation between successful uptake and retention, or learning. Mehmet’s case already showed that a learner will be unlikely to show signs of retention when evidence of CF uptake is lacking. Excerpt (9) below from Dinesh’s text provides further confirmation of this association between the two processes.

(9)  
S1: ... komen-3rd PL de *larve-3rd SG ... subject-verb agreement error  
...come the *larva ...  
[... the larvae come]  
S2: not applicable corrected utterance deleted  
S4: Eerst zijn-3rd PL ze *larve-3rd SG subject-verb agreement error  
First are they *larva  
[First they are larvae]

Whereas excerpts (3) and (4) showed that errors in noun pluralization such as the one in (9) are (partially) amenable to CF, Dinesh did not benefit from the correction targeting his erroneous plural marking on the noun larve (larva); he committed the same error both in his initial text (S1) and in his delayed post-test writing (S4). The reason behind this lack of retention might be found in the fact that Dinesh did not display any signs of uptake in relation to the feedback instance; instead of correcting or reformulating the S1 utterance in (9), Dinesh chose to delete it from his revised text (S2). We hypothesize that, although the direct CF Dinesh received had the potential to treat the pluralization error, it failed to do so because of the lack of uptake at S2, successful uptake having been proposed to be a “necessary step in developing longer term linguistic competence” (Ferris, 2004, p. 56).

In the same line of reasoning, we propose an explanation for Dinesh’s inability to improve his orthographical accuracy between the pre- and post-tests. Uptake of CF in the orthographical domain was relatively low; during revision (S2) Dinesh only succeeded in correcting 5 out of 10 erroneous uses and omissions of punctuation. This lack of successful
CHAPTER 5

CF uptake might account for the fact that Dinesh showed no accuracy gains over time in the orthographical domain.

Summary
Analysis of Dinesh’s sequential error rates seemed to show that he was relatively successful in incorporating the direct corrections he received during revision, but that this accuracy improvement was not retained in new pieces of writing. In-depth text analysis, however, revealed a different image of Dinesh’s accuracy development. We observed that (i) Dinesh deleted a considerable number of corrected utterances from his text during revision, and that, (ii) consequently, his level of successful CF uptake was relatively low; (iii) this limited degree of uptake translated into low levels of retention; and (iv) when CF was successfully taken up – as in the case of corrections targeting sentence level errors – Dinesh did show long-term accuracy improvement in relation to the provided CF.

5.5 Conclusions, discussion, and implications

It has been suggested that global measures of accuracy improvement (e.g. error rate comparison) might fail to give an adequate and complete picture of the effects of CF on L2 learners’ writing (Bruton, 2009a). Hence, we subjected the sequential writing performances of four L2 learners of Dutch to an in-depth accuracy analysis. In doing so the present multiple case-study aimed at (1) advancing our understanding on how and when written CF (fails to) affect(s) L2 learners’ accuracy development, and at (2) expanding insights into the amenability of different error types to direct and indirect CF.

Our findings proved that in-depth error analyses could indeed be considered a valuable supplement to the more common quantitative method of assessing accuracy development. We found that the detailed approach provided both different and additional information as compared to global accuracy measurement.

To begin with, the following three observations lead us to conclude that the effectiveness of CF is underestimated by overall error rate comparison: (i) For all four participants their progress, as measured by global scores only, was partially masked by the new errors they committed in their subsequent writing; around 50% of the errors in pupils’ (delayed) post-test texts was not related in any way to errors in their pre-test writing and the CF they received on those S1 problems. As Bruton (2009a) rightly argued, CF could not be expected to promote the accurate use of features it did not target. (ii) Moreover, Emre’s
writing showed that, in certain cases (cf. excerpt (2) and (3)), CF might lead to partial acquisition of a corrected feature. This type of improvement, however, will not be represented in global accuracy scores, since it is only the nature and not the number of errors that changes (Bruton, 2007; 2010). (iii) Finally, whereas in Mehmet’s and Dinesh’s cases overall error rate comparison seemed to show that CF harmed their accuracy development rather than promoting it, in-depth analyses revealed that – under the right circumstances – correction did trigger long-term accuracy improvement.

The detailed error analyses we performed, did not only yield different results from global accuracy measurement. As we aimed for, it also provided us with additional insights concerning the effectiveness of error correction. Regarding this study’s first objective – exploring when and how CF might be beneficial to L2 learners’ accuracy development – we were able to make three interesting observations.

First of all, our findings suggested that the level of successful CF uptake might be predictive of a pupil’s success in acquiring a target form; analyzing the texts written by Mehmet and Dinesh, we observed a lack of feedback retention in those cases where the pupils failed to uptake the provided corrections during revision. This finding is in line with the suggestion in the literature that successful uptake might play a facilitative role in acquisition (e.g. Chaudron, 1977; Lightbown, 1998; Loewen, 2004; Sheen, 2004; Swain, 1985), or could even be a prerequisite for learning (e.g. Ferris, 2004). The idea behind this theorized relation between uptake and retention is that noticing is a necessary condition for acquisition (Schmidt, 1995), and that uptake might be interpreted as (one of the) manifestation(s) of noticing.

A second observation pertaining to the relation between uptake and learning is that successful CF uptake is no guaranty for long-term acquisition of the targeted feature. As excerpt (7) from Mehmet’s work showed, CF targeting an exceptionally deep-rooted and potentially fossilized error might not result in retention, even though the correction was taken up during revision.

Thirdly, in-depth analyses of learners’ errors and corrections provided us with some hints on why pupils sometimes failed to take up the CF they were provided with. The majority of the corrections which did not lead to successful uptake targeted punctuation and capitalization errors. We argued that our learners might have failed to incorporate these feedback instances because of their low saliency; not only their small (physical) size, but also the fact that errors of these types do not greatly affect the comprehensibility of a text, make them relatively easy for learners to overlook. Moreover, we suggested that the level of successful CF uptake might be influenced by affective factors, such as the way in which
learners perceive a revision task (e.g. Storch & Wigglesworth, 2010). We found that Dinesh added and deleted whole utterances (including ones he received CF on), and hypothesized that his interpretation of the revision’s goal might have been broader than copying his initial text and incorporating the corrections as the instruction prescribed. By deleting corrected sentences, his (potential) level of CF uptake decreased.

The second aim of our study was to find out if all errors are equally amenable to direct and indirect CF. Firstly, our findings lead us to believe that – apart from errors which are too entrenched in a learner’s interlanguage grammar to be susceptible to CF (cf. excerpt (6), (7), and (8)) – all types of errors can benefit from correction; our participants showed to be able to gain accuracy on a broad range of linguistic features (e.g. determiner usage, agreement, inflection, punctuation, capitalization, and sentence structure), both on a local level (i.e. within one constituent) and a more global, sentential level. What is important to note, is that the improvement brought about by CF is not item-based; the case of Emre, for example, showed that the CF he received on the article omissions in his initial text led him to realize all obligatory determiners in his post-test writing, not just the ones that were corrected before.

While all investigated error types proved to be correctable, we did find that different types of errors vary in their responsiveness to indirect CF. We observed, for example, that the indirect CF Emre received led him to insert a determiner in front of each noun in his post-test writing, as Dutch grammar prescribes for definite contexts (cf. excerpt 2). However, the indirect corrections left him with insufficient clues to self-edit his idiosyncratic lexical error in excerpt (1). Excerpt (4) from Nathalie’s text, on the other hand, illustrated how direct CF did lead to accuracy development of an idiosyncratic feature. Based on these observations – and following Ferris (1999) – we proposed that, whereas rule-based errors seem to be suitable candidates for self-correction based on indirect CF, direct CF might be more beneficial to learners’ errors that are not rule-governed. This proposition is in line with DeKeyser’s (2003) suggestion that arbitrary form-function mappings will require more explicit learning processes. Although indirect CF is explicit in its corrective nature, it is less explicit than direct correction because it does not provide a learner with the target form. In those cases where a learner cannot depend on a clear rule when trying to deduce a target structure (e.g. when revising a lexical error), indirect CF might not be explicit enough to be beneficial. Therefore, maximally explicit feedback methodologies – such as direct CF –

Evidence of learning might be harder to find with respect to CF instances that are necessarily targeted at a specific item only, such as corrections of lexical and (most) spelling errors; chances are small that the corrected items can be used in a new text.
QUALITATIVE STUDY

could be expected to be most effective in developing the accurate use of idiosyncratic features.

Even though the suggestions in this chapter are based on the individual performances of only four L2 learners, and one should be cautious generalizing the findings, we believe the present research added to our understanding on how and when written CF is beneficial to accuracy development. The detailed analyses adopted in the study provided valuable insights into the relation between CF uptake and acquisition, and the amenability of different types of errors to direct and indirect error correction, which would have stayed hidden behind the numbers in a quantitative approach.
Chapter 6

Conclusion

6.1 Introduction

The role of corrective feedback (CF) in the process of second language acquisition (SLA) is a topic that is receiving a lot of research attention. Sheen (2010a) attributed this continuing research interest in CF to “the significance it carries for both SLA theory building and language pedagogy” (p. 177). The series of studies presented in this thesis were carried out with this dual perspective in mind. Framed within a cognitive account of SLA, they set out to investigate the potential of comprehensive written CF in fostering second language (L2) development, and opted to translate empirical insights into both theoretical and pedagogical implications.

The first two investigations reported on in this dissertation are two quantitative, tightly controlled classroom-based studies incorporating pre-test, treatment, and (delayed) post-test sessions (cf. Chapter 3, N = 66 and Chapter 4, N = 268). They explored the effects of CF and two control treatments on the accuracy development of secondary school pupils learning Dutch as their L2. The third study is a qualitative exploration into the efficacy of written error correction (cf. Chapter 5, N = 4). The writing performances of four L2 learners (who were selected from among the participants of the large-scale study presented in Chapter 4) were subjected to an in-depth examination to advance the understanding of how and when individual learners benefit from written CF.

In this concluding chapter, I will synthesize the main findings produced by the three empirical studies presented in Chapters 3, 4, and 5 (cf. section 6.2). In addition, I will discuss their theoretical and practical implications (cf. sections 6.3 and 6.4). Finally, I will discuss the limitations of the present work, and sketch out some open issues and directions for future research (cf. section 6.5).
6.2 Synthesis of main findings

The empirical work presented in this thesis aimed at contributing to the error correction debate (e.g. Ferris, 1999; 2004; Truscott, 1996; 1999; 2007) by addressing six major issues: (1) the value of comprehensive CF as an editing tool, (2) the value of comprehensive written CF for L2 development, (3) the differential effectiveness of direct and indirect error correction, (4) factors mediating CF efficacy, (5) the potential negative side-effects of written error correction, and (6) how and when individual learners (fail to) benefit from CF. This section will summarize and synthesize the insights that the three empirical studies (cf. Chapters 3, 4, and 5) offered with respect to these topics. Finally, I will discuss some findings from Chapter 5 that do not directly relate to the above mentioned issues, but which are nevertheless relevant, as they demonstrate the added value of in-depth accuracy measures.

6.2.1 The value of comprehensive CF as an editing tool

The first objective of the present empirical work was to investigate whether comprehensive or unfocused CF (i.e. correction of all existing errors) has the ability to help learners develop more effective revision and editing skills. As did earlier research (e.g. Ashwell, 2000; Fathman & Whalley, 1990; Ferris, 1997; Ferris & Roberts, 2001; Sachs & Polio, 2007), the studies presented in Chapters 3 and 4 proved that comprehensive CF is indeed a useful editing tool, by showing that CF helps learners to enhance the accuracy of an initial manuscript during revision. Importantly, findings showed that CF has an added value above revision as such. Whereas pupils who were asked to revise their texts on their own (i.e. without any available feedback) were able to arrive at more accurate revisions, their revised texts were still significantly less accurate than those of pupils who received CF. These findings show that CF is a useful tool that can help learners in becoming more successful writers, which is important from a learning-to-write perspective on L2 writing.

6.2.2 The value of comprehensive CF for L2 acquisition

However interesting from a learning-to-write point of view, the observation that CF leads to increased accuracy during revision is relatively insignificant from an SLA perspective. The crucial concern of SLA research is the long-term effect of pedagogical interventions such as CF, or L2 development. Evaluating the SLA potential of CF therefore necessarily involves
CONCLUSION

“a comparison between two independently written works” (Truscott & Hsu, 2008, p. 293), instead of contrasting an initial text and its revision.

Whereas recent research already revealed that learners receiving selective or focused CF (i.e. correction of errors within a specific category only) are able to use the targeted feature(s) more accurately in new pieces of writing (e.g. Bitchener, 2008; Bitchener & Knoch, 2008; Bitchener & Knoch, 2009; Bitchener & Knoch, 2010a; Bitchener & Knoch, 2010b; Ellis, Sheen, Murakami, & Takashima, 2008; Sheen, 2007; Sheen, 2010b), the studies presented in this thesis were the first to show that unfocused CF leads to learning. The investigations reported in Chapters 3 and 4 found that pupils who received comprehensive CF made fewer errors in new pieces of writing than learners whose errors had not been corrected. Moreover, this positive effect of CF proved to be durable; it was not only visible one week after pupils received the corrections (cf. Chapters 3 and 4) but also retained over a four week period (cf. Chapter 4). Besides that, in-depth exploration of learners’ accuracy development (cf. Chapter 5) showed that the improvement brought about by CF was not item-based. One of the case-study participants (i.e. Emre) omitted several articles in his initial text. The CF he received on those errors led him to realize all obligatory determiners in his post-test writing, not just the ones that were corrected before. Similarly, the CF Dinesh received on constituent omissions resulted in improvement beyond the accurate use of the specific corrected items; even though, obviously, he did not use the same sentences (i.e. items) he received feedback on in his post-test writing, Dinesh’s newly written texts did not contain any incomplete sentences.

6.2.3 The relative efficacy of direct and indirect CF

Apart from exploring the overall effectiveness of comprehensive CF, the present work aimed at investigating the relative value of different comprehensive CF methodologies for learners’ accuracy development. Feedback interventions have often been categorized as either direct (i.e. teacher indicates errors and target forms) or indirect CF (i.e. teacher only indicates errors, not the target forms) types. The main feature distinguishing these two feedback approaches is the level of learner involvement in the correction process. When receiving direct corrections, the only thing a learner is expected to do during revision, is incorporating the target forms as provided by the teacher. Indirect CF, on the other hand, requires more active engagement from the part of the student. Using indirect CF can be seen as a problem-solving activity in which learners need to correct their own errors. The teacher only indicates that an error has been made (e.g. by means of underlining, error
codes, etc). In the present studies indirect CF took the form of error coding (i.e. the provision of different codes representing different types of errors, see Appendix C).

The results from the small-scale study reported in Chapter 3, led to the cautious suggestion that direct CF might be more beneficial to accuracy development than indirect correction. This conclusion was based on the observation that, during revision, pupils who received direct feedback committed significantly fewer errors than their classmates receiving indirect CF. Moreover, the direct CF group outperformed pupils in the two control groups when writing a new text, while the indirect CF group did not. However, the post-test difference between the direct and indirect feedback groups only took the form of a trend at a p-value of .06.

In section 6.2.4, I will explain that it might not be feasible to come to a unidirectional conclusion on the relative merits of direct and indirect CF. The studies presented in Chapters 4 and 5 revealed that the efficacy of the two feedback methodologies is dependent on the type of error that is targeted.

**6.2.4 Factors mediating CF efficacy**

There are many factors that could be expected to mediate the efficacy of (a specific type of) CF. The benefits of error correction might be dependent, for example, on the nature of the targeted error, learners' attitudes towards the provided feedback, or the goal a teacher pursues by providing CF. The studies presented in this thesis explored the effects of three such potentially mediating factors: (1) error type, (2) learners' educational level, and (3) the topic of the task on which feedback is given.

**Error type**

It has been argued that different types of errors might not all be equally amenable to CF (e.g. Ferris, 1999; Truscott, 1996), because morphological, syntactic, and lexical errors represent gaps within different domains of linguistic knowledge (e.g. Schwartz, 1993). The studies presented in Chapters 4 and 5 therefore explored the effects of CF on errors within different domains. Whereas Chapter 4 contrasted grammatical and non-grammatical problems, Chapter 5 took a more in-depth look into the CF responsiveness of separate error types. Both studies furthermore investigated if the nature of a targeted error interacts with the efficacy of direct and indirect correction.
**CONCLUSION**

*Grammatical and non-grammatical errors*

Chapter 4 opted to test Truscott’s (2001; 2007) hypothesis that error correction might lead to improved accuracy of non-grammatical features, but that grammatical errors are insusceptible to CF. This study’s analyses therefore distinguished between grammar errors (e.g. morphosyntactic errors) and errors outside the grammatical domain (e.g. spelling errors). Results refuted Truscott’s claim by showing that CF enabled learners to improve both their grammatical and non-grammatical accuracy over time.

*Separate error types*

It could be expected that within the two broad domains distinguished in Chapter 4, separate error types still differ in their level of CF responsiveness (e.g. Ferris, 1999; Truscott, 2001). The multiple case-study presented in Chapter 5 therefore provided a more in-depth exploration of the amenability of different types of errors to written CF. The study showed that CF facilitated accuracy development on a broad range of linguistic features (e.g. determiner usage, agreement, sentence structure, inflection, and punctuation), both on a local level (i.e. within one constituent) and on a more global, sentential level. These findings led to the conclusion that syntactic, morphological, lexical, and orthographical errors all benefit from CF.

Results revealed one situation in which CF might be unable to lead to learning. One of the case-study participants did not show any sign of feedback retention when CF targeted one of the most frequent errors in his work (i.e. a common Dutch spelling problem). It was hypothesized that extremely persistent and potentially fossilized errors might be too deeply entrenched in a learner’s interlanguage system to be susceptible to CF.

*Interactions between error type and CF methodology*

Both Chapters 4 and 5 aimed at exploring if the relative efficacy of direct and indirect CF is dependent on the type of error that is targeted. Findings suggested that this is indeed the case. The study reported in Chapter 4 showed that only direct correction promoted grammatical accuracy development. Pupils’ non-grammatical accuracy, on the other hand, proved to benefit most from indirect CF; in the non-grammatical domain, the effect of indirect CF was more durable than that of direct CF.

Findings from the qualitative study presented in Chapter 5 brought to light that the effectiveness of the different CF methodologies and error type are even more intricately related. As Ferris (1999) predicted, the efficacy of indirect correction showed to be dependent on whether or not a targeted feature is rule-based in character. The case-studies
revealed that errors in the use of rule-based features – article omissions for example – are suitable candidates for self-correction based on indirect CF. However, indirect CF proved to be unsuccessful in remedying linguistic problems that are not rule-governed, such as lexical errors. It was argued that in those cases where a learner cannot depend on a clear rule when trying to deduce a target structure, indirect CF might not be explicit enough to be beneficial. The qualitative data also confirmed Ferris’ hypothesis that more explicit feedback types, such as direct CF, might be more effective in developing the accurate use of idiosyncratic features; one of the case-studies illustrated that direct CF did foster accuracy development of a non-rule-based feature.

At first sight, the findings from Chapters 4 and 5 seem to be contradictory with respect to the effect that direct and indirect CF have on lexical errors. Being categorized non-grammatical errors in Chapter 4, indirect correction was implicitly claimed to be most effective in addressing learners’ lexical problems. Conversely, Chapter 5 suggested that indirect CF provides learners with insufficient clues to self-edit their lexical errors because they are idiosyncratic in nature. A closer look at Chapter 4’s non-grammatical error category, and its amenability to direct and indirect CF, is needed in order to unravel this alleged disagreement in findings. When considering the distribution of errors across separate error types, it becomes clear that lexical problems only made up 13 percent of the total number of errors within Chapter 4’s broad non-grammatical error category. When subsequently removing the lexical errors from the measure of non-grammatical accuracy and rerunning the analyses on the delayed post-test data, another important observation can be made. The original delayed post-test analyses reported in Chapter 4 (cf. section 4.6.4) already showed that indirect correction was significantly more effective in remedying learners’ non-grammatical errors than writing practice or self-correction without CF. The difference between the direct and indirect treatments, however, just failed to reach significance (p = .06) in the initial analyses. Interestingly, this difference between the two CF methodologies (in favor of indirect CF) did become significant (p = .04) after the lexical errors were taken out of the non-grammatical error category. What the study in Chapter 4 thus showed, is that non-grammatical errors outside the lexical domain profit most from indirect correction. Hence, the new analyses were able to resolve the conflict in the findings from Chapters 4 and 5.

**Educational level**

The study reported in Chapter 4 furthermore investigated the influence of learners’ educational level on CF effectiveness. From a pedagogical perspective this is an interesting
issue, because for teachers it is useful to know if learners across different educational levels respond differently to (a specific type of) written CF.

Additionally, exploring the potentially mediating effect of learners' educational level on feedback efficacy might lead to theoretical implications. It has been claimed that factors such as learners’ levels of language proficiency and meta-linguistic awareness influence the degree to which pupils are able to benefit from error correction in general, and indirect CF in particular (e.g. Ferris, 2004; Hyland & Hyland, 2006). As explained in Chapter 2, pupils’ educational level was assumed to be indicative of their level of (meta-)linguistic competence.

Chapter 4 corroborated the expectation that the higher-level pupils outperformed the lower-level learners on the different linguistic measures used in the study (i.e. measures of written accuracy, lexical and structural complexity, and vocabulary knowledge). However, no significant interactions were found between the efficacy of the different CF treatments and learners’ level of educational. One possible interpretation of this finding is that it rebuts the hypothesis that CF is more beneficial to learners with higher levels of meta-linguistic awareness (e.g. Ferris, 2004; Hyland & Hyland, 2006; Sheen, 2007). However, as will be discussed in section 6.5.3, it might also be that the presumption that learners’ educational level is related to their level of (meta-)linguistic competence is not a valid one.

**Task topic**

Writing researchers have suggested that it is not just learners' level of L2 proficiency which determines the quality of the texts they produce. Other factors, such as a task’s topic, also proved to contribute to students’ writing performance (e.g. Meuffels & Van den Bergh, 2005; Schoonen, 2005). In this line of reasoning, it might be conceivable that learners’ knowledge about, interest in, and/or familiarity with the topic of the text they receive feedback on, also influences their motivation or ability to benefit from the provided CF. Chapter 3 therefore explored this potential mediating effect of a task’s topic by including writing tasks on two different themes. Results showed that the different topics of the two tasks used in the present study did not influence the extent to which learners benefited from the CF they received on their writing.

---

1 It needs to be noted that, in this type of studies, linguistic accuracy only constitutes one aspect of writing performance (which is most commonly assessed by means of holistic writing scores).
6.2.5 Potential harmful side-effects of CF

CF opponents proposed that, besides being ineffective, error correction could be expected to have a detrimental effect on learners’ accuracy development (e.g. Krashen, 1982; Truscott, 1996). The first argument raised to support this claim, is that correction and revision are too time-consuming, and that valuable class time could be spent more effectively on extra writing practice. Secondly, it was argued that the single goal of CF is making learners aware of the errors they committed. This awareness was thought to lead learners to avoid making the same errors in future writing, and therefore to result in less complex language use. The present research explored these two potential harmful effects of written CF by (1) comparing the performance of pupils receiving CF to that of learners who were given an extra opportunity to practice their writing skills, and by (2) assessing the effect of CF on the complexity of learners’ writing.

The results presented in this thesis refuted both of the above mentioned claims. The findings in Chapters 3 and 4 showed that pupils who received CF outperformed learners who were allowed an extra opportunity to practice their writing skills. Moreover, the results reported in Chapter 4 proved that CF did not lead to avoidance; CF did not result in a reduction of the lexical and/or structural complexity of learners’ writing.

6.2.6 How and when individual learners (fail to) benefit from CF

Chapter 5 subjected the sequential writing performances of four L2 learners of Dutch to an in-depth accuracy analysis to be able to gain insights into how and when written CF affects individual L2 learners’ accuracy development. Even though only based on the performances of four pupils, and caution should be taken in generalizing the findings, this study was able to lead to two interesting observations.

First, findings suggested that the potential of CF in yielding a learning effect might be dependent on a learner’s level of CF uptake (where uptake was operationalized as the successful revision of an erroneous utterance based on CF). The case-studies showed that in the majority of cases where pupils committed errors in their post-test writing on which they already received feedback, this lack of CF retention could be traced back to a lack of CF uptake during revision.

Secondly, qualitative analyses of learners’ errors and corrections provided some clues on why pupils sometimes fail to take up the CF they are provided with. It was suggested in Chapter 5 that CF instances targeting errors which are low salient – either in terms of their physical size or because they do not greatly affect the comprehensibility of a text – might be relatively easy to overlook, and will therefore not always be taken up.
Moreover, the case-studies seemed to suggest that affective factors, such as the way in which learners perceive a revision task (e.g. Storch & Wigglesworth, 2010), might influence learners’ level of CF uptake.

6.2.7 The value of in-depth accuracy measures

It has been suggested that global measures of accuracy improvement (e.g. error rate comparison) might fail to give an adequate and complete picture of the effects of CF on L2 learners’ accuracy development (Bruton, 2009a; Storch, 2010). The findings reported in Chapter 5 proved that this is indeed the case, and that in-depth error analyses could be considered a valuable supplement to the more common quantitative method of assessing CF effectiveness. Besides providing additional insights into the correctability of different error types and learners’ engagement with CF (as was explained in sections 6.2.4 and 6.2.6), the detailed approach also led to the conclusion that the effectiveness of CF is underestimated by overall error rate comparison. This conclusion was based on the following observations. First, global scores were shown to partially mask learners’ progress by including errors in new texts that were not related in any way to the errors corrected in the initial texts. As Bruton (2009a) rightly argued, CF could not be expected to promote the accurate use of features it did not target. Secondly, in some cases, learners only showed retention for part of a feedback instance. CF targeting article omissions, for example, led one of the case-study participants (i.e. Emre) to realize the obligatory determiners in new pieces of writing, but not the correct ones (i.e. Dutch differentiates between neuter and non-neuter determiners). This type of partial acquisition of a corrected feature, however, is not represented in global accuracy scores because such measures adopt all-or-nothing criteria (i.e. accurate/inaccurate) (Bruton, 2007; 2010).

6.3 Theoretical implications

This section evaluates how the findings presented in this thesis add to the theoretical understanding of the role of written CF in SLA. It will successively discuss (1) the value of CF to accuracy development, (2) implications relating to the role of attention in CF processing, (3) the relation between CF uptake and interlanguage development, and (4) the level of CF explicitness that is required for SLA.
6.3.1 CF and accuracy development

As was explained in Chapter 2, one’s conceptualization of (the relation between) implicit and explicit L2 knowledge and the value attributed to explicit knowledge as such, are reflected in one’s expectations with respect to the ability of CF to aid SLA. Whereas implicit L2 knowledge could be defined as knowledge of the language, explicit knowledge constitutes of knowledge about the L2 system.

Scholars opposing the usefulness of CF (e.g. Krashen, 1985; Truscott, 1996) have claimed that, at best, error correction may lead to the development of explicit knowledge, and that such knowledge about the language is of very little value in itself. The position that explicit knowledge will never become implicit, then leads those CF contestants to conclude that CF has no role in L2 development.

The more common view, however, is that both implicit and explicit L2 knowledge are important links in the SLA process (e.g. DeKeyser, 1998; Doughty & Williams, 1998; Hulstijn, 1995; Hulstijn & Schmidt, 1994; Long & Robinson, 1998; McLaughlin, 1990; Schmidt, 1990; Schmidt & Frota, 1986; Swain, 1985), and that CF can be beneficial to interlanguage development, either by assisting the proceduralization of explicit knowledge (e.g. DeKeyser, 1998) or by fostering cognitive processes such as noticing (the gap) and hypothesis testing (e.g. Swain, 1991).

Whereas it is still an empirical question whether learners store CF as implicit or explicit knowledge (Bitchener & Knoch, 2010a), the present research (together with other recent written CF studies) does counter the claim that CF can only lead to ‘pseudolearning’ (Truscott, 1996). The fact that learners still benefited from the CF they received when writing new texts, demonstrates that error correction has the potential to lead to accuracy development. Importantly, these CF invoked accuracy gains proved to be durable; Chapter 4 showed that the observed positive effects of written error correction were retained over a four-week period, and therefore refutes Truscott’s (1996) idea that CF can only lead to a transient form of L2 knowledge. What makes the present findings even more promising is that the studies reported on in this thesis used realistic, communicative writing tasks which invited learners to focus on content rather than on language as an object. As pointed out by several researchers (e.g. Long, 2007; Ellis, 2010), such free constructed responses arguably afford the most valid measure of language development.

---

2 Bitchener & Knoch (2010a) were even able to show that the beneficial effects of written CF were retained as long as ten months after the feedback had been provided.
6.3.2 CF and attention

Attention is a crucial notion within cognitive SLA accounts. It is generally believed that only the subset of the input that is attended to is available for further processing (e.g. Robinson, 2003; Schmidt, 1990; 1994; 2001; Sharwood Smith, 1993; VanPatten & Cadierno, 1993). One of the hypothesized essential properties of attentional resources is that they are limited (e.g. Schmidt, 2001). Two related implications that follow from the present work concern the attentional demands associated with comprehensive CF and learners’ allocation of attentional resources to written accuracy and complexity.

Attentional demands of comprehensive CF

Advocates of a focused CF approach have suggested that targeting a (couple of) specific error type(s) should be expected to be more effective than comprehensive or unfocused error correction (e.g. Bitchener, 2008; Ellis et al., 2008; Sheen, 2007). In their view, comprehensive CF might fail to facilitate SLA because L2 learners have a limited processing capacity. Those in favor of focused correction argue that asking learners to deal with CF targeting a broad range of linguistic features at the same time might lead to a cognitive overload, and thus prohibit feedback processing.

In showing that comprehensive CF has the potential to yield sustained accuracy improvement, the present findings do not support the hypothesis that processing unfocused CF is too cognitively demanding. Instead, they suggest that learners have enough attentional resources available to attend to many different linguistic features within one text. As proposed in Chapter 2, it might be the offline character of writing that prevents learners from becoming cognitively overloaded when presented with unfocused corrections (e.g. Sheen, 2010a). What is important to note, however, is that the texts pupils wrote in the present studies were rather short (i.e. around 120 words). It is an empirical question if comprehensive CF is also an effective means of addressing longer texts.

Attention, accuracy, and complexity

Limited capacity models of attention hypothesize that focusing learners’ attention on accuracy (i.e. by providing CF) will lead to a reduction of the linguistic complexity of learner output (e.g. Skehan, 1998; Skehan & Foster, 2001). Those models assume that accuracy and complexity draw upon a single pool of attention and are therefore in constant competition for attentional resources.

The present findings are not in line with the accuracy/complexity trade-off predicted by single-resource accounts of attention. Chapter 4 showed that learners were able to
improve the accuracy of their writing without sacrificing on the end of structural or lexical complexity3. It could therefore be suggested that these outcomes support a multiple-resource perspective on attention (e.g. Robinson, 2003; 2005). Within such a model, accuracy and complexity are not presumed to be in competition because these two performance dimensions of linguistic form are thought to be closely connected.

An alternative interpretation of the findings is also possible. It is feasible that it was only during text revision that pupils’ attention was explicitly focused on accuracy, since that was the moment in which they were made aware of their non-target-like language use (i.e. by means of CF). During revision, all attentional resources could be allocated to accuracy because all other features (i.e. content, complexity) had already been taken care of. When writing new texts, on the other hand, pupils’ attention was not explicitly drawn to accuracy issues. The fact that the writing tasks used in the present studies were communicative in nature without any inherent focus on language form, might have allowed learners to allocate minimal attention to accuracy in the post-test sessions, thus leaving sufficient attentional capacity to spend on the linguistic complexity of their writing. In this line of reasoning, one would not expect any accuracy/complexity trade-off effects, because there was never a need to pay focal attention to both accuracy and complexity simultaneously.

Unrelated to attentional capacity issues, but connected to the interrelationship between accuracy and complexity, is Truscott’s (1996) claim that CF may lead to avoidance of complex language use. He stated that by inducing learner stress and anxiety of committing the same errors in future writing, CF could make learners avoid the erroneous constructions when writing a new text. By demonstrating that CF did not lead to simplified writing, the findings presented in Chapter 4 rebut Truscott’s claim.

6.3.3 CF uptake and SLA

Uptake can be defined as a learner’s response to a CF instance or, as Lyster and Ranta (1997) put it, uptake is “what the student attempts to do with the teacher’s feedback” (p. 49). Uptake is successful when the original error is repaired. Although successful uptake in itself does not constitute evidence of interlanguage development (e.g. Ellis et al., 2001; Long, 2007), several researchers have proposed that a learner’s level of CF uptake might be predictive of L2 acquisition (e.g. Ellis & Sheen, 2006; Lightbown, 1998; Loewen, 2004; Lyster, 1998; Sheen, 2004). In their view, the fact that a learner was able to come to a

3 It has to be noted that the study reported on in Chapter 4 only included one measure of lexical complexity (i.e. Guiraud’s index) and one measure of structural complexity (i.e. subordination index).
target-like reformulation based on CF gives reason to believe that noticing has taken place, which, in turn, facilitates SLA (e.g. Schmidt, 1990; 2001).

The qualitative investigation presented in Chapter 5 suggests that – at least for the four learners in this study – CF uptake was indeed a good predictor of the success in acquiring a target form. As observed in both Chapter 4 and Chapter 5, in general, CF was able to trigger sustained gains in the accurate use of a targeted feature. When CF failed to lead to accuracy development, however, this lack of feedback retention could be traced back to a lack of successful CF uptake during revision. Interestingly, the observations reported in Chapter 5 also demonstrated that CF uptake forms no guaranty for long-term acquisition. It was shown that CF targeting an exceptionally deep-rooted and potentially fossilized error might not result in retention, even when the correction is taken up during revision. This finding thus validates the claim that the level of successful uptake, although predictive of acquisition, was not a reliable developmental measure in itself (e.g. Ellis, Basturkmen, & Loewen, 2001; Long, 2007).

6.3.4 The required level of CF explicitness
As was explained in Chapter 2, researchers have advanced theoretical arguments in favor of both direct and indirect CF methodologies. For example, indirect correction has been claimed to have a greater potential to foster SLA than direct CF because indirect CF engages learners in a more profound form of language processing (e.g. Ferris, 1995; Lalande, 1982). On the other hand, direct CF could be argued to be superior to indirect feedback forms because it presents learners with the kind of explicit information that is needed for cognitive learning processes, such as noticing and hypothesis testing (e.g. Bitchener & Knoch, 2010b). The studies discussed in Chapters 4 and 5 of this dissertation provide support for both of these hypotheses, and suggest that the required level of CF explicitness is dependent on the type of error targeted.

Chapter 4 showed that the beneficial effects of the less explicit, cognitively demanding indirect approach are more durable than those of direct correction when CF targets linguistic features that pupils are able and confident to self-correct, that is their non-grammatical errors (e.g. spelling errors, punctuation errors). However, this study revealed that feedback needs to be more explicit (i.e. direct) in order to lead to the development of grammatical (i.e. morphosyntactic) accuracy.

The qualitative study presented in Chapter 5 furthermore revealed that the needed level of CF explicitness depends on the idiosyncrasy of the targeted error. Whereas rule-based errors (e.g. inflection errors) showed to be suitable candidates for self-correction
based on indirect CF, direct CF proved to be more beneficial to learners’ errors in features that are not rule-governed (e.g. lexical errors). This finding is in line with DeKeyser’s (2003) suggestion that arbitrary form-function mappings will require more explicit learning processes. In those cases where a learner cannot depend on a clear rule when trying to deduce a target structure indirect CF might not be explicit enough to be beneficial. Therefore, maximally explicit feedback methodologies – such as direct CF – could be expected to be most effective in developing the accurate use of idiosyncratic features.

The observation that different error types show differential levels of responsiveness to direct and indirect CF methodologies, is in line with the claim that no single form of CF can be effective in addressing all linguistic error types (e.g. Ferris, 1999; Truscott, 1996), because morphological, syntactic, and lexical errors represent gaps within different domains of linguistic knowledge (e.g. Schwartz, 1993). It is important to note, however, that all error types have proven to be correctable, and that Truscott’s (2001; 2007) hypothesis that grammatical errors are insusceptible to CF could thus be rejected.

### 6.4 Pedagogical implications

Even though it is important to keep in mind that it is unsure if the findings of the present experiments would hold in a real-world class situation (cf. section 6.5), it is still worth considering some pedagogical implications from the studies presented in this thesis. This section discusses (1) the value of CF for language learning, (2) the efficacy of authentic, comprehensive correction, (3) the importance of revision activities, (4) the complementary value of different CF methodologies, and (5) the potential of CF within content-focused contexts.

#### 6.4.1 The value of written CF for language learning

The main and foremost suggestion that follows from the present findings (as well as from other recent written CF studies) is that CF is a useful instrument that L2 teachers can employ to help learners improve their written accuracy. CF has shown to be valuable both as an editing tool and as a SLA intervention; it does not only enable students to improve a particular piece of writing during revision, but learners still benefit from CF when writing a new text. Just one feedback treatment already proved to be effective in constituting long-term accuracy gains. These beneficial effects could only be expected to be greater if learners are offered CF on additional occasions.
6.4.2 The efficacy and ecological validity of comprehensive CF

A second important pedagogical contribution of the empirical studies reported in this thesis, is that they revealed that comprehensive written CF has the potential to yield a learning effect. Until now, the only robust evidence on the long-term effects of written CF came from studies exploring the efficacy of focused correction, that is CF targeting one specific type of errors (e.g. Bitchener, 2008; Bitchener & Knoch, 2008; Bitchener & Knoch, 2009; Bitchener & Knoch, 2010a; Bitchener & Knoch, 2010b; Ellis et al., 2008; Sheen, 2007; Sheen, 2010b). However important these findings may be for SLA theory, some scholars have questioned their practical applicability (e.g. Ferris, 2010; Storch, 2010). They pointed out that, when providing CF, teachers usually opt to improve the overall accuracy of their students’ writing, not just the use of one specific linguistic feature. This might be particularly true in contexts where the main pedagogical focus is on communicating content rather than on language as an object (e.g. Anderson, 2010). It has therefore been claimed that comprehensive CF is a more authentic feedback methodology (e.g. Anderson, 2010; Ferris, 2010; Hartshorn et al., 2010; Storch, 2010). By exploring the efficacy of comprehensive CF, the present studies – even though still experimental in set-up – tried to reflect real classroom conditions in terms of the type of feedback provided. The finding that an authentic CF methodology such as comprehensive error correction leads to L2 acquisition could thus be considered of great relevance to teachers.

6.4.3 The importance of revising opportunity

A third implication that can be drawn from the work at hand is that teachers should allow learners the opportunity to revise their texts based on the provided feedback. Asking students to perform revision activities could be expected to foster SLA because “producing the correct form may help learners automatize their L2 production” (Loewen, 2004, p. 157). Moreover, producing accurate revisions could be considered a manifestation of pushed output, and should therefore be expected to promote L2 acquisition by triggering noticing (the gap) and hypothesis testing (Swain, 1985; 2005). Chapter 5 provided evidence of these potential advantages of text revision by showing that CF instances were more likely to lead to improved accuracy in new texts when learners were able to use them effectively during revision (i.e. when they showed signs of CF uptake).

In relation to the above, it could be considered important to make students aware of the goals and value of CF provision and revision activities, and to stimulate learners’ motivation to engage with the provided feedback. One of the case-studies reported in Chapter 5 showed, for example, that the potential beneficial roles of CF-based revision are
CHAPTER 6

not served when learners misperceive its goals. Other researchers have also attested that affective factors – such as attitudes towards (a certain type of) CF, perceived goals, motivation – may influence the success of CF interventions (e.g. Bruton, 2009b; Goldstein, 2006; Hyland, 1998; Storch, 2010; Storch & Wigglesworth, 2010; Swain & Lapkin, 2003).

6.4.4 The complementary value of different CF methodologies

As was discussed in sections 6.2.4 and 6.3.4, the extent to which learners are able to benefit from direct and indirect CF methodologies, might be dependent on the type of error targeted. It was argued that explicit forms of CF, such as direct correction, are more effective in solving complex (e.g. morphosyntactic) or idiosyncratic (e.g. lexical) errors, while indirect CF might lead to more durable accuracy improvement when targeting relatively simple features (e.g. punctuation). Moreover, Ferris (2010) proposed that direct and indirect feedback methodologies might serve different goals. When the focus is on improving learners’ L2 composition skills, and CF is intended to help learners in becoming self-employed writers, indirect CF methods might be considered more useful because they demand a more active form of learner engagement. When mainly opting for language learning, on the other hand, direct correction might be judged the most beneficial approach because it provides the kind of efficient and explicit input necessary for L2 acquisition. I conclude that, depending on teachers’ goals and the types of errors that need to be attended to, direct and indirect CF interventions can complement each other.

The same holds for focused (or selective) and unfocused (or comprehensive) correction strategies. Since both CF types have now been shown to be effective, teachers can choose which approach they think is most appropriate in a specific situation. It might be useful, for example, to alternate comprehensive CF methodologies with intensive, focused correction when targeting linguistic features that have proven to be persistently problematic for a given learner.

6.4.5 The CF potential within a content-focused context

The present empirical work has shown that providing written CF is an effective means of drawing learners’ attention to language form in a content-focused context (i.e. biology class). One of the implications that follows is that time spent on correction of written products could be considered well-invested, even when learners’ L2 development is not the primary concern of the context the CF is provided in.

Albeit only based on anecdotic evidence, I would claim that this finding is important because content teachers might often be reluctant to devote time and attention to language
related issues; even in the context of the present research, where schools’ policy prescribed a language sensitive approach to content teaching (e.g. Van Eerde & Hajer, 2008), many teachers felt that providing language-related feedback was the task and responsibility of the Dutch language teacher⁴. Additionally, some of them failed to see how the provision of CF would benefit their teaching, and stated that the language sensitive character of their pedagogical approach mainly involved extra glossing of content-specific vocabulary⁵.

However, research has shown that the language problems L2 learners have to deal with, go deeper than lacking the appropriate content-related vocabulary (e.g. Prenger, 2005; Van den Boer, 2003), and signaled a need for attention to linguistic form and accuracy in content-focused contexts (Schooten & Emmelot, 2004). The studies reported in this dissertation showed that written CF can be one of the tools content-teachers can use to fulfill this need.

One more argument in favor of providing CF within content-focused contexts and not just in language classrooms (within the same educational program), resides in the notion of transfer-appropriate learning (Segalowitz, 1997; 2000). It could be expected that feedback provided in a content-focused class is more beneficial to pupils’ future accuracy performance in this same context than CF provided in a fully language oriented lesson; what is learned from processing feedback in a certain context is conceivably more transfer-appropriate than the type of knowledge gained from CF provided in another setting.

6.5 Limitations and directions for further research

Although the studies presented in this thesis provide clear evidence in favor of written comprehensive CF, there are a number of limitations to the present empirical work that need to be acknowledged, and which could be considered in future research. This section

---

⁴ The present study did not incorporate any planned or structured interviews with teachers about their attitudes towards providing CF within their own classroom context. The claim that content teachers might be hesitant to use CF is based on informal conversations between the researcher and the teachers.

⁵ Taken together, these observations might lead one to conclude that the schools participating in the present research may not have fully implemented the language sensitive approach. A comparable observation was already done by Swain and Carrol (1987, as cited in Lyster, 2007), who noticed that, in immersion classrooms, form and function were kept surprisingly distinct even though the pedagogical goal was to teach language through content.
consecutively discusses (1) the specific context of the work at hand, (2) its scope (i.e. L2 accuracy development), and (3) some methodological issues.

### 6.5.1 Research context

It has been pointed out, that it is important for research to acknowledge its contextualization (e.g. Manchón, 2009; Ortega, 2009). In the case of the present work, this is of particular significance, because its setting is very different from the contexts other written CF studies were conducted in. On the one hand, this could be considered a strength of the investigations at hand; several scholars have called for research within diverse contexts (e.g. Manchón, 2009; Ortega, 2009; Révész, 2007). On the other hand, it is not self-evident that findings across contexts are readily comparable, or that findings from one setting can automatically be transferred to any other context.

The present research context differs from that of other recent studies into the efficacy of written CF (e.g. Bitchener, 2008; Bitchener & Knoch, 2008; Bitchener & Knoch, 2009; Bitchener & Knoch, 2010a; Ellis et al., 2008; Sheen, 2007; Sheen, 2010b; Truscott & Hsu, 2008) with respect to the age of the learners, the language under investigation, and learners’ L2 proficiency level. Whereas in other CF studies, participants were generally adults with an intermediate level of L2 English proficiency, the population in the present research consisted of adolescent, highly proficient L2 learners of Dutch.

Additionally, while earlier CF work tested CF efficacy within instructed SLA (i.e. at university level), the educational setting under investigation in the present book could be characterized as an (early) immersion context. As was explained in Chapters 1, 3, and 4, the secondary schools participating in the research project adopted a language sensitive approach to content teaching (e.g. Van Eerde & Hajer, 2005). This pedagogical approach opts to integrate content and language instruction, in order to cater for the special needs of L2 learners. Where the L2 is the main object of study in instructed SLA settings, the first and foremost focus of the present instructional context is on content (e.g. biology).

One conclusion from the present work, that might be dependent on the context it was derived from, is that only direct CF leads to grammatical accuracy development. It needs to be acknowledged that the participants in the studies at hand received a limited

---

6 The research contexts of some of the early CF studies (e.g. Kepner, 1991; Lalande, 1982) differed from the present setting on an additional factor. They investigated the effect of CF within foreign language (FL) instruction rather than in L2 contexts.

7 Participants had an average of 10 years experience in the target language; most of them started learning Dutch in school at age four and were 14 years old at the time of data collection.
amount of explicit grammar instruction. They could therefore be expected to have low levels of meta-linguistic knowledge and awareness. As previously suggested in Chapter 4, it might well be that learners had too little knowledge about the L2 to be able to really and fully benefit from indirect CF, and that another research context would have led to different conclusions on the differential value of direct and indirect CF. It would thus be interesting to replicate the studies presented in this thesis in other pedagogical settings and with other populations, to test the generalizability of the present findings.

6.5.2 Research scope
As was explained in Chapter 2, the focus of the studies presented in this dissertation was exclusively on the effects of written CF on learners’ accuracy development (i.e. the outcome of the language learning process). As a result, the present work only enabled further understanding of CF’s potential in yielding a learning effect, in terms of accuracy gains. It did not provide any direct insights, however, into the SLA processes – such as noticing (the gap), hypothesis testing, problem-solving – that may underlie CF efficacy (e.g. Adams, 2003; Santos et al., 2010; Storch & Wigglesworth, 2010; Swain & Lapkin, 2002; Qi & Lapkin, 2001), and/or the affective factors that may explain learners’ (lack of) feedback uptake (e.g. Bruton, 2010; Goldstein, 2006; Hyland, 1998; Storch & Wigglesworth, 2010; Swain & Lapkin, 2003).

In particular, the findings from the qualitative study presented in Chapter 5 would have been more illuminating when an attempt would have been made to tap into learner internal processes and/or learners’ perceptions of, attitudes towards, and levels of motivation associated with the different CF treatments. In order to gain a full understanding of the role of CF in SLA, future research that triangulates product data, process data (e.g. think-aloud protocols, stimulated recall protocols), and socio-cultural data (e.g. retrospective interviews, questionnaires) would thus be warranted.

6.5.3 Research methodology
Even though the present empirical work tried to address the design-related problems of earlier written CF studies (e.g. Chandler, 2003; Kepner 1991; Polio et al., 1998; Semke, 1984; see Chapter 3 for a review), the present methodological set-up still had its flaws.

To begin with, the studies reported in Chapters 3, 4, and 5 did not include immediate post-tests out of practical considerations and time constraints. Instead, the first post-test was only administered one week after the treatment session (i.e. the moment of feedback provision). Inclusion of an immediate post-test would have enabled the comparison of
immediate and delayed CF effects, and could have provided insights into the proportion of initial accuracy gains (i.e. as evident in the performance on an immediate post-test) that are retained over a longer period of time (i.e. as evident in the performance on a post-test and delayed post-test).

The second shortcoming also relates to the number of post-tests used in the present work. Although the study presented in Chapter 4 already proved that the accuracy gains brought about by comprehensive written CF were retained over a one-month period, it would have been interesting to further investigate the durability of comprehensive CF effects, by means of adding (a) more delayed post-test(s).

A final drawback that needs to be acknowledged, concerns the operationalization of learners’ meta-linguistic awareness in Chapter 4. Learners’ educational level was presumed to be indicative of their level of meta-linguistic awareness. The rationale behind this assumed relationship was twofold (cf. Chapter 2, section 2.6.3). First, it was suggested that pupils from different educational strands could be anticipated to vary in (among other things) their level of meta-linguistic awareness, because a considerable part of the assessment that is used in the Netherlands to place pupils within different levels of secondary education concerns learners’ (meta-)linguistic skills (Cito, 2010). Secondly, it was noted that the Dutch national framework of reference for language skills sets different goals for pupils in different levels of secondary education with respect to language proficiency and meta-linguistic knowledge (Expertgroep Doorlopende Leerlijnen Taal en Rekenen, 2008). However, the study reported in Chapter 4 did not test if pupils with a higher level of education indeed disposed of a greater awareness of the L2 system. Without empirical evidence on the validity of its operationalization, any firm claims on the mediating effect of learners’ meta-linguistic awareness on CF efficacy would be unwarranted. Hence, the conclusion that Chapter 4’s results do not constitute evidence for the hypothesis that indirect CF is more helpful for learners with higher levels of meta-linguistic awareness (e.g. Ferris, 2004; Hyland and Hyland, 2006), is a very tentative one.

### 6.6 Concluding remarks

Whereas the value of written CF for L2 acquisition has been heavily contested (e.g. Truscott, 1996; 1999; 2007), the studies presented in this thesis (as well as other recent CF studies) provide robust evidence on the efficacy of error correction in L2 writing. The fact that the accuracy improvement brought about by written CF was shown to be durable,
rebuts Truscott’s (1996) claim that correction can only lead to a superficial and transient type of L2 knowledge. I conclude that, by offering learners opportunities to notice the gaps in their developing L2 systems, test interlanguage hypotheses, and engage in metalinguistic reflection, written CF has the ability to foster SLA and to lead to accuracy development. The present empirical work thus advances the theoretical understanding of the language learning potential of written CF, and shows that comprehensive CF is a useful pedagogical tool.


REFERENCES


Bruton, A. (2009b). Improving accuracy is not the only reason for writing, and even if it were... System 37(4), 600-613.


REFERENCES


REFERENCES


REFERENCES

language related problems and linguistic needs of first year pupils in secondary education, academic year 2002-2003.] Groningen: Etoc, University of Groningen.


Han, Z-H. (2002). Rethinking the role of corrective feedback in communicative language teaching. RELC Journal 33(1), 1-34.


REFERENCES


Meuffels, B., & Van den Bergh, H. (2005). De ene tekst is de andere niet, the language as-a-fixed-effect fallacy revisited: Methodologische implicaties. [Not every text is the same, the language as-a-fixed-effect fallacy revisited: Methodological implications.] Tijdschrift voor Taalbeheersing 27(2), 106-125.


[Standardization of reference levels in primary education.] Amsterdam: SCO-Kohnstamm Institute, University of Amsterdam.


REFERENCES


REFERENCES


REFERENCES


REFERENCES

Mathematics Education (pp. 1215-1225). Sant Feliu de Guíxols: Fundemi IQS - Universitat Ramon Llull.


Appendices

Appendix A: Writing tasks

Metamorphosis tasks

The metamorphosis of a butterfly – original Dutch version (p. 160)
This task was used for all conditions in the pre-test sessions (S1) of the studies presented in Chapters 3 to 5. Logically, this task was also used for the direct CF, indirect CF, and self-correction conditions in the treatment sessions (S2) of those studies.

The metamorphosis of a butterfly – English version (p. 162)
English translation of the original Dutch writing task.

The metamorphosis of a honeybee – original Dutch version (p. 164)
This task was used for the practice condition in the treatment session (S2) of the large-scale study (cf. Chapter 4).

The metamorphosis of a ladybug – original Dutch version (p. 166)
This task was used for all conditions in the post-test sessions (S3) of the studies presented in Chapters 3 to 5.

The metamorphosis of a wasp – original Dutch version (p. 168)
This task was used for the practice condition in the treatment session (S2) of the small-scale study reported on in Chapter 3, and for all conditions in the delayed post-test session (S4) of the studies presented in Chapters 4 and 5.

All tasks are included in their original Dutch versions, and one task (i.e. the metamorphosis of a butterfly) has been translated into English as an example.
Van ei tot vlinder

Hieronder vind je een schrijfopdracht. Lees eerst de opdracht goed door. Begin daarna pas met schrijven.

Opdracht: Schrijf een e-mail aan een klasgenoot.

Volgende week heb je een biologieproefwerk. Eén van de onderwerpen waarover je leraar vragen zal stellen, is de metamorfose van een vlinder. Je beste vriend(in) was ziek toen de leraar vertelde hoe er uit een eitje uiteindelijk een vlinder groeit. Hij/zij heeft je daarom gevraagd of je in een e-mail kunt uitleggen wat er op de plaatjes op de plastic kaart gebeurt. Schrijf deze e-mail aan je vriend(in). Beschrijf voor iedere stap van de metamorfose wat er gebeurt of verandert.

Bij de beoordeling van je tekst letten we er op of je alle stappen duidelijk hebt beschreven. Verder is het belangrijk dat je hele zinnen gebruikt. Je e-mail moet minstens 15 regels lang zijn!

SUCCES!

Naam: ………………………………………………….
School: ………………………………………………. 
Klas: …………………………………………………
Leeftijd: ………………………………………………

Code: …………………………………………………
Sessie: S1
Begintijd: ……………………………………………
Eindtijd: ……………………………………………
Metamorfose van een vlinder

De ontwikkeling van ei tot vlinder

1. Eitje
2. Larve
3. Kleine rups
4. Volwassen rups
5. Vervellen
6. Pop
7. Vlinder
From an egg into a butterfly

Below, you will find a writing assignment. Only start writing when you have read the assignment carefully.

Task: Write an e-mail to a classmate.

You will have a biology test next week. One of the topics the teacher will question you about is the metamorphosis of a butterfly. Your best friend was ill when the teacher explained how an egg will eventually evolve into a butterfly. Your friend therefore asked you to write him/her an e-mail to explain what happens in the pictures on the plastic card.

Write this e-mail to your friend. Explain for every step of the metamorphosis what is happening or changing.

When evaluating your text, we will assess if you succeeded in clearly describing the different steps of the metamorphosis process. Moreover, it is important to write in whole sentences, Your e-mail has to be at least 15 lines in length!

GOOD LUCK!
Metamorphosis of a butterfly
The transformation from egg to butterfly

Egg

Larva

Small caterpillar

Full-grown caterpillar

To shed skin

Pupa

Butterfly
Van ei tot bij

Hieronder vind je een schrijfopdracht. Lees eerst de opdracht goed door. Begin daarna pas met schrijven.

Opdracht: Schrijf een e-mail aan een klasgenoot.

Volgende week heb je een biologie proefwerk. één van de onderwerpen waarover je leraar vragen zal stellen, is de metamorfose van een bij. Je beste vriend(in) was ziek toen de leraar vertelde hoe er uit een eitje uiteindelijk een bij groeit. Hij/zij heeft je daarom gevraagd of je in een e-mail kunt uitleggen wat er op de plaatjes op de plastic kaart gebeurt.

Schrijf deze e-mail aan je vriend(in). Beschrijf voor iedere stap van de metamorfose wat er gebeurt of verandert.

Bij de beoordeling van je tekst letten we er op of je alle stappen duidelijk hebt beschreven. Verder is het belangrijk dat je hele zinnen gebruikt. Je e-mail moet minstens 15 regels lang zijn!

SUCCES!
Metamorfose van een bij
*De ontwikkeling van ei tot bij*
Van ei tot lieveheersbeestje

Hieronder vind je een schrijfopdracht. Lees eerst de opdracht goed door. Begin daarna pas met schrijven.

Opdracht: Schrijf een e-mail aan een klasgenoot.

Volgende week heb je een biologieproefwerk. Eén van de onderwerpen waarover je leraar vragen zal stellen, is de metamorfose van een lieveheersbeestje. Je beste vriend(in) was ziek toen de leraar vertelde hoe er uit een eitje uiteindelijk een lieveheersbeestje groeit. Hij/zij heeft je daarom gevraagd of je in een e-mail kunt uitleggen wat er op de plaatjes op de plastic kaart gebeurt.

Schrijf deze e-mail aan je vriend(in). Beschrijf voor iedere stap van de metamorfose wat er gebeurt of verandert.

Bij de beoordeling van je tekst letten we er op of je alle stappen duidelijk hebt beschreven. Verder is het belangrijk dat je hele zinnen gebruikt. Je e-mail moet minstens 15 regels lang zijn!

SUCCES!
Metamorfose van een lieveheersbeestje
De ontwikkeling van ei tot lieveheersbeestje

Eitje

Larve
Larve
Larve

Vervellen

Pop

Jong lieveheersbeestje
Volwassen lieveheersbeestje
Hieronder vind je een schrijfopdracht. Lees eerst de opdracht goed door. Begin daarna pas met schrijven.

Opdracht: Schrijf een e-mail aan een klasgenoot.

Volgende week heb je een biologieproefwerk. Eén van de onderwerpen waarover je leraar vragen zal stellen, is de metamorfose van een wesp. Je beste vriend(in) was ziek toen de leraar vertelde hoe er uit een eitje uiteindelijk een wesp groeit. Hij/zij heeft je daarom gevraagd of je in een e-mail kunt uitleggen wat er op de plaatjes op de plastic kaart gebeurt.

Schrijf deze e-mail aan je vriend(in). Beschrijf voor iedere stap van de metamorfose wat er gebeurt of verandert.

Bij de beoordeling van je tekst letten we er op of je alle stappen duidelijk hebt beschreven. Verder is het belangrijk dat je hele zinnen gebruikt. Je e-mail moet minstens 15 regels lang zijn!

SUCCES!
Metamorfose van een wesp
De ontwikkeling van ei tot wesp

1. Eitje
2. Larve
3. Pop
4. Pop
5. Pop
6. Pop
7. Jonge wesp
8. Volwassen wesp
Laundry instructions tasks

**How to wash a T-shirt (p. 171)**
This task was used for all conditions in the pre-test session (S1) of the study presented in Chapter 3. Logically, this task was also used for the direct CF, indirect CF, and self-correction conditions in the treatment session (S2) of this study.

**How to wash a pair of jeans (p. 173)**
This task was used for the practice condition in the treatment session (S2) of the study presented in Chapter 3.

**How to wash a sweater (p. 175)**
This task was used for all conditions in the post-test session (S3) of the study presented in Chapter 3.
Hoe was je een T-shirt?

Hieronder vind je een schrijfopdracht. Lees eerst de opdracht goed door. Begin daarna pas met schrijven.

Opdracht: Schrijf een e-mail aan een vriendin.


Maar je vriend(in) weet niet wat al die plaatjes betekenen. Gelukkig heb jij net in de biologieles geleerd wat er met de verschillende wassymbolen bedoeld wordt. Je belooft je vriend(in) een e-mail te schrijven waarin je uitlegt hoe hij/zij het T-shirt moet wassen.

Schrijf deze e-mail aan je vriend(in). Beschrijf voor ieder symbool wat het betekent en leg stap voor stap uit wat hij/zij moet doen om het T-shirt weer helemaal schoon en netjes te krijgen. Leg ook uit wat hij/zij niet mag doen en waarom niet. Je kunt de symbolen opzoeken in het schema op de plastic kaart.

Bij de beoordeling van je tekst letten we er op of je voor alle symbolen hebt uitgelegd wat ze betekenen en duidelijk hebt gemaakt wat je vriend(in) met het T-shirt moet doen. Verder is het belangrijk dat je hele zinnen gebruikt.

Je e-mail moet minstens 15 regels lang zijn!

SUCCES!
## Wassymbolen

<table>
<thead>
<tr>
<th>Wassymbolen</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Wassen</strong></td>
</tr>
<tr>
<td>30 graden was</td>
</tr>
<tr>
<td>40 graden was</td>
</tr>
<tr>
<td>40 graden was koelwaterstand</td>
</tr>
<tr>
<td>40 graden was warmwaterstand</td>
</tr>
<tr>
<td><strong>Niet wassen</strong></td>
</tr>
<tr>
<td>60 graden was koelwaterstand</td>
</tr>
<tr>
<td>60 graden was warmwaterstand</td>
</tr>
<tr>
<td>95 graden was warmwaterstand</td>
</tr>
</tbody>
</table>

- **Bleken**
- **Koud bleken**
- **Niet bleken**

- **Kan in droogtrommel**
- **Hete voeglessen textiel**
- **Normaal textiel**
- **Niet in de droogtrommel**

- **Strijken**
- **Laag strijken**
- **Warm strijken**
- **Heet strijken**
- **Niet strijken**

- **Chemisch reinigen**
- **Gewone reiniging**

- **Niet chemisch reinigen**
- **Speciale reiniging**
Hoe was je een spijkerbroek?

Hieronder vind je een schrijfopdracht. Lees eerst de opdracht goed door. Begin daarna pas met schrijven.

Opdracht: Schrijf een e-mail aan een vriendin.

De ouders van je beste vriend(in) zijn op vakantie. Hij/zij is nu alleen thuis en moet zelf het huishouden doen. Zijn/haar spijkerbroek, die hij/zij heel graag morgen aan wil, is vies. Die moet hij/zij nu dus ook zelf wassen. Maar hoe moet dat? Je vriend(in) heeft gezien dat er in de broek een label zit met allerlei symbolen. Het label ziet er zo uit:

Maar je vriend(in) weet niet wat al die plaatjes betekenen. Gelukkig heb jij net in de biologieles geleerd wat er met de verschillende wassymbolen bedoeld wordt. Je belooft je vriend(in) een e-mail te schrijven waarin je uitlegt hoe hij/zij de spijkerbroek moet wassen.

Schrijf deze e-mail aan je vriend(in). Beschrijf voor ieder symbool wat het betekent en leg stap voor stap uit wat hij/zij moet doen om de spijkerbroek weer helemaal schoon en netjes te krijgen. Leg ook uit wat hij/zij niet mag doen en waarom niet. Je kunt de symbolen opzoeken in het schema op de plastic kaart.

Bij de beoordeling van je tekst letten we erop of je voor alle symbolen hebt uitgelegd wat ze betekenen en duidelijk hebt gemaakt wat je vriend(in) met de spijkerbroek moet doen. Verder is het belangrijk dat je hele zinnen gebruikt. Je e-mail moet minstens 15 regels lang zijn!

SUCCES!
Hoe was je een trui?

Hieronder vind je een schrijfopdracht. Lees eerst de opdracht goed door. Begin daarna pas met schrijven.

Opdracht: Schrijf een e-mail aan een vriendin.

De ouders van je beste vriend(in) zijn op vakantie. Hij/zij is nu alleen thuis en moet zelf het huishouden doen. Zij/haar lievelingstrui, die hij/zij heel graag morgen aan wil, is vies. Die moet hij/zij nu dus ook zelf wassen. Maar hoe moet dat? Je vriend(in) heeft gezien dat er in de trui een label zit met allerlei symbolen. Het label ziet er zo uit:

![Label met symbolen](image)

Maar je vriend(in) weet niet wat al die plaatjes betekenen. Gelukkig heb jij net in de biologieles geleerd wat er met de verschillende wasymbolen bedoeld wordt. Je belooft je vriend(in) een e-mail te schrijven waarin je uitlegt hoe hij/zij de trui moet wassen.

Schrijf deze e-mail aan je vriend(in). Beschrijf voor ieder symbool wat het betekent en leg stap voor stap uit wat hij/zij moet doen om de trui weer helemaal schoon en netjes te krijgen. Leg ook uit wat hij/zij niet mag doen en waarom niet. Je kunt de symbolen opzoeken in het schema op de plastic kaart.

Bij de beoordeling van je tekst letten we erop of je voor alle symbolen hebt uitgelegd wat ze betekenen en duidelijk hebt gemaakt wat je vriend(in) moet doen. Verder is het belangrijk dat je hele zinnen gebruikt. Je e-mail moet minstens 15 regels lang zijn!

SUCGES!
Wassymbolen
Appendix B: Instructions treatment session

Instructions on the use of direct corrective feedback

Instructie fouten verbeteren

Voor je liggen de twee teksten die je vorige week geschreven hebt. De teksten zijn nu nagekeken en gecorrigeerd.

Probeer nu je eigen teksten te verbeteren met behulp van de correcties die je in je teksten ziet. Schrijf de verbeterde teksten op de twee bladzijden hierachter.

Als je klaar bent met verbeteren, steek je je vinger op.

Vergeet niet je naam onderaan dit blad te schrijven!

Succes!

Instruction error correction

In front of you are the two texts that you wrote last week. They have now been corrected.

Try to revise your own texts based on the provided corrections. Write down the revised versions on the two pages following this one.

When you are done, please let us know.

Don’t forget to fill in your name at the bottom of this page!

Good luck!

---

1 Both the original Dutch instructions and English translations are provided.
Instructions on the use of indirect corrective feedback

Instructie fouten verbeteren met behulp van codes

Voor je liggen de twee teksten die je vorige week geschreven hebt. De teksten zijn nu nagekeken en gecorrigeerd. Maar misschien niet op de manier die je gewend bent.

In je teksten zie je verschillende codes. Die codes geven aan dat je een fout gemaakt hebt. In de codelijst die je hebt gekregen, kun je zien wat de verschillende codes betekenen. Bekijk de codes en voorbeelden op de lijst nu eerst eens rustig.

Probeer daarna je eigen teksten te verbeteren met behulp van de codes die je in je teksten ziet. Schrijf de verbeterde teksten op de twee bladzijden hierachter.

Als je klaar bent met verbeteren, steek je je vinger op.

Vergeet niet je naam onderaan dit blad te schrijven!

Succes!

Instruction error correction using error codes

In front of you are the two texts that you wrote last week. They have now been corrected, but maybe not in the way you are used to.

In your text you will encounter different codes. These codes signal that you have committed an error. The hand-out you received tells you what the different codes mean. Start by studying the codes and examples on your hand-out.

When you are done, try to revise your own texts using the error codes that you encounter. Write down the revised versions on the two pages following this one.

When you are done, please let us know.

Don’t forget to fill in your name at the bottom of this page!

Good luck!
**Instructions on self-correction**

*Instructie teksten verbeteren*

Voor je liggen de twee teksten die je vorige week hebt geschreven. Lees deze teksten nu nog eens aandachtig en serieus door. Heb je alles duidelijk en uitgebreid genoeg opgeschreven? Heb je geen spelfouten gemaakt? Heb je goede, hele zinnen gebruikt?

Probeer nu je eigen teksten te verbeteren. Schrijf de verbeterde teksten op de twee bladzijden hierachter. (Als je niets wilt verbeteren, moet je toch de teksten overschrijven.)

Als je klaar bent met verbeteren, steek je je vinger op.

_Vergeet niet je naam onderaan dit blad te schrijven!_

_Succes!_

**Instruction error correction using error codes**

*In front of you are the two texts that you wrote last week. They have now been corrected, but maybe not in the way you are used to.*

_In your text you will encounter different codes. These codes signal that you have committed an error. The hand-out you received tells you what the different codes mean. Start by studying the codes and examples on your hand-out.*

_When you are done, try to revise your own texts using the error codes that you encounter. Write down the revised versions on the two pages following this one._

_When you are done, please let us know._

_Don’t forget to fill in your name at the bottom of this page!_

_Good luck!_
## Appendix C: Error coding system (indirect CF)

<table>
<thead>
<tr>
<th>Code</th>
<th>Betekenis (meaning)</th>
</tr>
</thead>
</table>
| ( )  | **Verkeerde vorm van het woord (incorrect word form)**  
   | *Voorbeeld:* Hij heeft een nieuw boek (gekrijgen).  
   | *Verbetering:* Hij heeft een nieuw boek *gekregen*.  
   |   | *Voorbeeld:* De man *(word)* een beetje te dik.  
   | *Verbetering:* De man *wordt* een beetje te dik.  
   |   | *Voorbeeld:* De man *(loopt)* gisteren op straat.  
   | *Verbetering:* De man *liep* gisteren op straat.  
   |   | *Voorbeeld:* Hij heeft een *(nieuwe)* boek gekregen.  
   | *Verbetering:* Hij heeft een *nieuw* boek gekregen.  |
|     | **Verkeerd woord gebruikt (incorrect word choice)**  
   | *Voorbeeld:* Het *man* loopt op straat.  
   | *Verbetering:* De *man* loopt op straat.  |
| S   | **Spelfout (spelling error)**  
   | *Voorbeeld:* Ik ga om 9 uur naar *sgool*S.  
   | *Verbetering:* Ik ga om 9 uur naar *school*.  |
| *...* | **De woordvolgorde klopt niet (incorrect word order)**  
   | *Voorbeeld:* Ik heb *gekocht een boek*.  
   | *Verbetering:* Ik heb *een boek gekocht*.  |
| V   | **Een woord vergeten (omission)**  
   | *Voorbeeld:* Hij heeft *V* nieuw boek gekocht.  
   | *Verbetering:* Hij heeft *een* nieuw boek gekocht.  
   |   | *Voorbeeld:* *V V* Gisteren een nieuw boek gekocht.  
   | *Verbetering:* *Ik heb* gisteren een nieuw boek gekocht.  |
| \  | **Een woord teveel gebruikt (addition)**  
   | *Voorbeeld:* Hij gaat /naar in\ huis.  
   | *Verbetering:* Hij gaat *naar* huis.  |
Een punt, komma, dubbele punt of vraagteken vergeten of fout gebruikt (interpunctiefout) *(punctuation error)*

Voorbeeld: Hij gaat naar huis i
Verbetering: Hij gaat naar huis.

Voorbeeld: Hij gaat naar huis, want hij is ziek,i hij heeft hoofdpijn.
Verbetering: Hij gaat naar huis, want hij is ziek. Hij heeft hoofdpijn.

Een hoofdletter vergeten *(capitalization error)*

Voorbeeld: Het is 5 uur. zij gaat naar huis.
Verbetering: Het is 5 uur. Zij gaat naar huis.

Geen hele zin gebruikt *(incomplete sentence)*

Voorbeeld: Zeep gebruiken.
Verbetering: Je moet zeep gebruiken.
Hallo,

Hieronder stellen we je een aantal vragen over de talen die je spreekt. We willen graag weten welke taal of talen je gebruikt in verschillende situaties. Je antwoorden worden anoniem verwerkt en alleen gebruikt voor het onderzoek.

1. Naam: ........................................................................................................
2. Leeftijd: ............ jaar  geslacht man □  vrouw □
3. School: ................................................................. Klas: ...........
4. Mijn geboorteland is: ...........................................................................
5. Het geboorteland van mijn moeder is: ...................................................
   Het geboorteland van mijn vader is: .....................................................
6. Als je niet in Nederland bent geboren, hoe lang ben je dan in Nederland?
   ............. jaar ........... maanden
7. In welke landen heb je hoe lang gewoond?
   ................................ jaar in ......................................................
   ................................ jaar in ......................................................
   ................................ jaar in ......................................................
8. Welke talen spreek je? Begin met de taal die je het beste spreekt en eindig met de taal die je het minst goed beheerst.
   Taal ....................................  Taal ....................................
   Taal ....................................  Taal ....................................
   Taal ....................................
9. Hoe vaak spreek je deze talen? Begin met de taal die je het vaakst spreekt en eindig met de taal die je het minst vaak gebruikt.
   Taal ....................................
   Taal ....................................
   Taal ....................................
   Taal ....................................
   Taal ....................................
   • Op school ...........................................................................................
   • Met je ouders .................................................................................
   • Met je broers en zussen .................................................................
   • Met je oma en oma ...........................................................................
   • Met je vrienden/vriendinnen ...........................................................
   • In andere situaties, namelijk ...........................................................
   • Je spreekt je ....................................................................................

   • Op televisie kijk ik naar programma's in het ..................................
   • Ik lees boeken in het .................................................................
   • Ik lees tijdschriften in het .........................................................
   • Ik bekijk internetsites in het ........................................................

Hieronder heb je ruimte voor opmerkingen.

.......................................................................................................................
.......................................................................................................................
.......................................................................................................................
.......................................................................................................................
.......................................................................................................................
.......................................................................................................................
.......................................................................................................................
.......................................................................................................................

Dank je wel voor het invullen van de vragenlijst!
Appendix E: Vocabulary test

Instruction

INSTRUCTIE

Deze toets bestaat uit 108 zinnen. In elke zin is één woord onderstreept. Kies uit de vier antwoorden steeds de betekenis die bij het onderstreepte woord past. Als je denkt dat je het goede antwoord gevonden hebt, zet je een cirkeltje om dat antwoord.
Als je het antwoord echt niet weet, kies je het vijfde antwoord ‘Ik weet het echt niet’.

Een voorbeeld:

Er bleef een graat in mijn keel steken.
   a. botje van een kip
   b. botje van een vis
   c. stukje brood
   d. soort knoop
   e. Ik weet het echt niet.

Een graat is een botje van een vis, dus antwoord (b) is goed. Je zet dan een cirkeltje om antwoord (b).

Beantwoord de opgaven in de volgorde zoals ze in de toets staan. Je maakt dus eerst opgave 1, dan 2, 3, 4, 5, enz. Sla geen opgaven over.
Je hebt 30 minuten om de toets te maken. In deze 30 minuten moet je proberen zoveel mogelijk opgaven te maken, maar de toets hoeft aan het eind niet af te zijn.

Vergeet niet je naam, school en klas hieronder in te vullen!

Succes!

Naam: ………………………………………
School: ……………………………………..
Klas: ……………………………………….
Code: ………………………………………
Excerpt vocabulary test\(^1\)

Item 1 to 12, out of 108 total

1. Het lijkt me wat **voorbarig** om zoiets te zeggen.
   a. goedkoop
   b. onbehoorlijk
   c. te vroeg
   d. onvriendelijk
   e. Ik weet het echt niet.

2. Hij is leraar van **professie**.
   a. Hij geeft les en is professor.
   b. Hij is heel goed in zijn werk.
   c. Hij geeft les aan professoren.
   d. Zijn beroep is leraar.
   e. Ik weet het echt niet.

3. Dat zou ik anders **inschatten**.
   a. beoordelen
   b. aankondigen
   c. opschrijven
   d. doen
   e. Ik weet het echt niet.

4. Mag ik jouw kam **even lenen**?
   a. ding waarmee je schrijft
   b. ding waarmee je rekent
   c. ding waarmee je het eten snijdt
   d. ding waarmee je je haren netjes maakt
   e. Ik weet het echt niet.

5. Ze heeft **oogschaduw** op.
   a. Ze heeft een doekje voor haar oog.
   b. Ze heeft een blauw oog.
   c. Ze heeft haar ogen mooi gemaakt.
   d. Ze heeft een bril op tegen de zon.
   e. Ik weet het echt niet.

6. In dat land heerst **cholera**.
   a. Het gaat in dat land slecht met de economie.
   b. Het heeft in dat land al jaren niet geregeld.
   c. Veel mensen in dat land hebben een bepaalde ziekte.
   d. Men voert al jaren oorlog in dat land.
   e. Ik weet het echt niet.

7. Die kunstenaar werkt veel met **keramiek**.
   a. een soort ijzer
   b. een soort hout
   c. een soort steen
   d. een soort papier
   e. Ik weet het echt niet.

8. Wat een lelijke hoest heb jij!
   a. manier waarop je je haar draagt
   b. je gedrag in het openbaar
   c. ding dat je met je keel beschermt als je verkouden bent
   d. geluid dat je met je keel maakt
   e. Ik weet het echt niet.

9. Zij is een **bolleboos**.
   a. Zij drinkt erg veel.
   b. Zij doet erg veel aan sport.
   c. Zij kan erg goed leren.
   d. Zij vindt zichzelf erg mooi.
   e. Ik weet het echt niet.

10. Hij **spaart** voor een auto.
    a. Hij stopt voor een auto langs de weg.
    b. Hij is onder een auto gekomen.
    c. Hij verzamelt geld om een auto te kopen.
    d. Hij verzamelt foto’s van auto’s.
    e. Ik weet het echt niet.

11. Hij is lid van de **schutterij**.
    a. vereniging die dieren beschermert
    b. vereniging die de natuur beschermert
    c. muziekvereniging
    d. schietvereniging
    e. Ik weet het echt niet.

12. Tijdens het **spitsuur** is het druk.
    a. tijd waarop mensen van en naar hun werk gaan
    b. tijd waarop de dokter mensen ontvangt
    c. einde van de wedstrijd
    d. begin van de vakantie
    e. Ik weet het echt niet.

---

\(^1\) Adapted version of a receptive vocabulary test, called the Hazenberg & Hulstijn test, originally designed by Hazenberg (1994) (cf. Chapter 3).
Summary

This dissertation investigates the effects of corrective feedback on second language learners’ written accuracy in Dutch secondary education.

Chapter 1: Introduction

Chapter 1 draws up the context in which the studies presented in this thesis were carried out, and introduces the central concepts in this book.

For many pupils in Dutch secondary education the language of instruction (i.e. Dutch) is not their mother tongue. Research has shown that these second language (L2) learners lag behind their native peers in school success (e.g. Gijsberts & Herweijer, 2009). One of the explanations for this overall arrears in school performance might be that these L2 learners lack the Dutch repertoire that is necessary to meet the high linguistic demands they are faced with in school. This makes that attention to language is of the utmost importance, not only in language-oriented classes but also in classes whose overriding focus is on content (e.g. biology, geography). Many multilingual schools therefore adopt a language-sensitive approach to their content teaching (e.g. Hajer & Meestringa, 2004). Among the spearheads of this approach are the necessity to provide learners with ample opportunity to engage in productive language use, and the need for (oral and written) language-related feedback.

One of the crucial questions then is what this language-related feedback should look like. A commonly used feedback type targeting linguistic problems is error correction or corrective feedback (CF). CF can be defined as “any indication to the learner that his or her use of the target language is incorrect” (Lightbown & Spada, 2006, p. 197). Even though CF is widely used in language classrooms, its usefulness has been the topic of considerable debate in the field of second language acquisition (SLA). Some researchers even expect CF to harm language development rather than promoting it (e.g. Truscott, 1996). The central questions in this thesis are if, how, and when written CF positively affects the accuracy development of L2 learners.
Written CF can take many different forms (cf. table 1.1, p. 4). The CF methodologies examined in this book, are a direct and an indirect type of unfocused or comprehensive CF. An unfocused or comprehensive CF approach involves correction of all errors in a learner’s text, irrespective of their error category. This methodology contrasts with focused or selective CF approaches, in which a teacher chooses to target a (limited number of) specific linguistic feature(s) only (e.g. errors in article usage). The distinction between direct and indirect CF relates to the level of learner involvement in the correction process. When providing direct CF, the teacher identifies both the error and the corresponding target form (cf. example 1). Indirect CF, on the other hand, only indicates that an error has been made, and leaves it to the learner to derive the target form. There are different ways of supplying indirect CF; errors could be underlined or tallied, for example. In the present research, errors were coded (cf. example 2, and Appendix C).

<table>
<thead>
<tr>
<th>Example 1: directe correctieve feedback</th>
<th>Example 2: indirecte correctieve feedback</th>
</tr>
</thead>
<tbody>
<tr>
<td>Je moet <em>het</em> trui niet <em>chemisch</em> reinigen.</td>
<td>Je moet <em>het</em> trui niet chemisch<em>S</em> reinigen.</td>
</tr>
<tr>
<td>de <em>chemisch</em></td>
<td>( _ = wrong word, S = spelling error)</td>
</tr>
<tr>
<td>[You should not dry clean the sweater.]</td>
<td>[You should not dry clean the sweater.]</td>
</tr>
</tbody>
</table>

The studies that are reported in Chapters 3, 4, and 5 investigate the effectiveness of direct and indirect comprehensive CF by assessing the effects of written CF on the linguistic accuracy of learners’ output. Accuracy was defined as the target-like use of language forms (i.e. grammar, lexis, and orthography). In more concrete words, this dissertation aims at answering the following six questions:

7. Does comprehensive CF lead to improved accuracy in revised texts?
8. Does comprehensive CF lead to improved accuracy in new texts?
9. What CF methodology is most effective: direct or indirect CF?
10. What factors mediate the effectiveness of comprehensive CF?
11. Does written CF come with any negative side-effects that harm accuracy development?
12. How and when do individual learners (fail to) benefit from comprehensive CF?
Chapter 2: Theoretical and empirical background
Taking a cognitive perspective on SLA, Chapter 2 presents the theoretical and empirical background of the research reported in this thesis. The first part of the chapter explains why, from a theoretical point of view, it is to be expected that CF promotes L2 development. Then it reviews some of the objections against the use of CF, raised by CF opponents. Next, the chapter discusses several controversies with respect to the relative efficacy of different CF methodologies, and the CF amenability of different types of errors. Finally, a review of earlier research on CF effectiveness leads up to an overview of the issues investigated in this book.

Having access to plenty L2 input (e.g. Krashen, 1982) and being granted ample opportunity to engage in language production (e.g. Swain, 1985) are necessary conditions for successful SLA. However, research has shown that even when these conditions are met, L2 learners fail to acquire a native-like level of accuracy in their production (e.g. Swain, 1991). To be able to progress towards well-formedness in their L2, learners not only need language input and production, but also specific attention to linguistic form (focus-on-form). One way of directing learners’ attention to form-related aspects of their L2 output, is by providing them with CF (e.g. Ellis, 2005). Feedback can make learners realize where their own output is not yet target-like (noticing the gap) (e.g. Han, 2002).

Those opposing the use of CF in L2 classrooms, state that CF is unable to positively affect the SLA process. One of the arguments they put forward is that CF could at best result in explicit knowledge about the language system, but that correction will never lead to the acquisition of the type of implicit knowledge that enables learners to communicate in their L2 (e.g. Krashen, 1982; Truscott, 1996). Moreover, CF contestants claim that CF might harm L2 development. First of all, because correcting and revising takes up valuable time which could be spent more effectively on productive aspects of L2 instruction. Secondly, CF might lead learners to simplify their writing by avoiding errors that were corrected before.

Apart from the question if CF promotes SLA, issues such as which CF methodology is most effective, or if all error types are equally responsive to CF, are also debated. Various predictions have been made, for instance, with respect to the relative effectiveness of comprehensive vs. selective CF, and direct vs. indirect correction, and Truscott (1996) has claimed that CF could only be beneficial for non-grammatical errors (e.g. spelling errors), but never for errors within the grammatical domain (e.g. errors in sentence structure).

Earlier research has already revealed that CF, irrespective of its methodology, is effective in promoting accuracy improvement during revision. The more important and interesting question, however, is if CF yields a learning effect: does the feedback one
receives on a particular piece of writing positively affect the accuracy of newly written texts? Recent studies showed that this is indeed the case, when a selective approach to CF was adopted. Research on the potential long-term effects of comprehensive CF, on the other hand, suffered from methodological problems, and produced conflicting findings. The same holds for studies comparing the efficacy of direct and indirect CF. Moreover, issues such as the potential harmful side-effects of error correction, or factors possibly mediating CF effectiveness (e.g. the type of error targeted, learners’ educational level, task topic), have not yet been investigated (systematically) enough. The three empirical studies presented in this thesis set out to contribute to the settlement of these open questions.

Chapter 3: Small-scale study
Chapter 3 describes a small-scale study (N = 66) investigating the effects of direct and indirect comprehensive feedback.

The research reported in this chapter was conducted at two Dutch multilingual secondary schools that adopted a language-sensitive approach to content instruction. The participants were 66 pupils in their second year of secondary pre-vocational education. Pupils were randomly assigned to four different treatment groups: two experimental treatments and two control treatments. Pupils in the experimental groups received either direct or indirect CF on all errors in their texts. The errors committed by learners in the control groups were not corrected. These pupils revised their own writing without having access to any feedback (self-correction), or wrote two new texts when the other groups revised their initial writing products (practice). Including two control treatments served, among other things, the following two goals: (i) Comparing the effectiveness of CF to that of self-correction, made it possible to set apart effects of error correction from effects of the revision process as such. (ii) Contrasting the effects of CF to the effects of extra writing practice, allowed for testing Truscott’s (1996) hypothesis that the opportunity for additional production is more beneficial than correcting and revising.

The experiment included three sessions. In the first session (week 1 – pre-test), all pupils were asked to write two texts, on two different topics (cf. Appendix A). In the second session (week 2 – treatment session), pupils revised their initial texts based on the CF they received, revised their texts without any available feedback, or practiced their writing skills by writing two new texts. In the third session (week 3 – post-test), all learners – irrespective of the treatment group they were assigned to – again wrote two texts. To be able to measure the efficacy of the different treatments, the accuracy level of each text was
calculated by means of error ratio’s (i.e. (number of linguistic errors/total number of words) x 10).

The results of this first study revealed that comprehensive CF leads to accuracy improvement in L2 writing. Both pupils whose errors were corrected directly, and learners who received indirect CF, made fewer errors during the treatment session than pupils in the two control groups. Moreover, direct CF yielded a learning effect: the post-test texts of learners in the direct group contained significantly fewer errors than the post-test writing of pupils in the practice and self-correction groups. Hence, these findings do not support Truscott’s (1996) supposition that writing practice is more beneficial to accuracy development than correction and revision. To conclude, task topic proved not to influence the extent to which learners profit from written CF.

Chapter 4: Large-scale study
Chapter 4 presents the findings of a large-scale study (N = 268) into the effectiveness of comprehensive written CF.

Four Dutch secondary schools with multilingual student populations participated in the research project. The design of this study was largely similar to the methodology of the small-scale study as described in Chapter 3. However, the following five methodological changes were implemented: (i) Since the first study demonstrated that a task’s topic does not mediate the efficacy of CF, this variable was not included in the large-scale study. (ii) To be able to explore the influence of educational level on CF effectiveness, pupils from two strands of secondary education participated in this study: pupils in their second year of secondary pre-vocational education (i.e. vmbo-tl), and learners in their second year of higher general secondary education (i.e. havo). (iii) The large-scale study included four experimental sessions instead of three. A delayed post-test was administered four weeks after the treatment session to gain insights into the durability of CF effects. (iv) The study reported in Chapter 4 not only explored the effect of comprehensive CF on overall accuracy. To test Truscott’s (2001; 2007) claim that non-grammatical errors might be more correctable than errors in grammar, the overall accuracy measure was broken down into a measure of grammatical accuracy and a measure of non-grammatical accuracy. (v) Finally, the lexical diversity and structural complexity of pupils’ writing were also measured, to be able to determine if CF results in avoidance and simplification of language production.

Results from the large-scale study once again showed that comprehensive CF positively affects the accuracy of both revised and new texts. Moreover, the effect of CF
proved to be durable; even a month after learners received feedback, pupils in the CF groups still outperformed pupils in the control groups. Another important finding relates to the correctability of grammatical and non-grammatical errors. Whereas both error types proved to be amenable to CF, they were shown to differ in their responsiveness to direct and indirect correction. Only direct CF promoted pupils’ grammatical accuracy development. The effect of CF on learners’ non-grammatical accuracy, on the other hand, proved to be more durable when errors were corrected indirectly. Pupils’ educational level turned out not to influence CF efficacy. Lastly, no evidence was found of any detrimental side-effects of CF. The results reported in Chapter 4 proved that CF did not lead learners to avoid more complex language use. Moreover, pupils who received CF outperformed learners who were allowed an extra opportunity to practice their writing skills in all sessions. Hence, it is not the case that the time and energy invested in CF would have been better spent on additional writing practice.

Chapter 5: Qualitative study
Chapter 5 reports a multiple case-study (N = 4) into the effects of direct and indirect comprehensive CF.

The studies in Chapters 3 and 4 make an important contribution to the error correction debate by showing that comprehensive CF can promote L2 development. However, quantitative studies like these ones leave some interesting issues unanswered. First of all, quantitative CF research fails to provide insights into how and when individual learners benefit from error correction (Storch & Wigglesworth, 2010). Secondly, whereas the study in Chapter 4 revealed that both non-grammatical and grammatical errors are amenable to correction, it still remains unclear how separate error types within these two broad domains react to CF. Finally, Bruton (2009a, 2010) suggests that global accuracy measures (such as the error ratio’s in Chapters 3 and 4) might not provide an accurate and complete picture of the accuracy gains brought about by CF. One of the arguments he puts forward is that, in adopting all-or-nothing criteria (i.e. accurate/inaccurate), global measures will inevitably fail to quantify partial learning.

Taking the above mentioned reservations towards global accuracy measurement at heart, the study presented in Chapter 5 adopted qualitative, in-depth analyses to explore the effectiveness of direct and indirect comprehensive CF. For the multiple case-study, four pupils were selected from among the participants of the large-scale study described in Chapter 4: two learners from the direct CF group, and two pupils whose errors had been
corrected indirectly. To gain a better understanding of how and when learners (fail to) benefit from correction, pupils were selected who seemingly reacted very differently to the CF they received: for each feedback type, one participant’s error rate decreased after he had received CF, whereas the other pupil’s error rate increased between pre-test and post-tests.

The four writing products (i.e. pre-test text, revised version of pre-test text, post-test text, and delayed post-test text) of each learner were subjected to a detailed comparison. When a feedback instance led to a correct revision or a target-like reformulation, this was taken as evidence of successful feedback uptake. The target-like usage of a corrected feature in a new piece of writing was seen as proof of CF retention or accuracy development. Errors in learners’ post-test texts that did not bear any relation to the errors pupils committed during the pre-test, were classified new errors. To be able to explore the amenability of separate error types to CF, errors in pupils’ writing were classified into four broad domains (morphosyntax, lexicon, orthography, and pragmatics), each of which was subsequently subdivided into a number of separate error types (cf. table 5.1, p. 98).

The qualitative approach adopted in Chapter 5 provided both different and additional insights as compared to the quantitative studies reported in Chapters 3 and 4. To begin with, the effectiveness of CF proved to be underestimated by global accuracy measurement. Learners’ progress was partially masked by the new errors pupils committed during the post-tests. Moreover, in some cases, CF led to partial acquisition of a corrected feature. As Bruton (2010) suggested, this type of improvement is not represented in global accuracy scores. Additionally, global accuracy measures seemed to indicate that, for two of the four learners in the study, CF harmed their accuracy development rather than promoting it. However, in-depth analyses revealed that – under the right circumstances – these learners also benefited from CF.

Moreover, the multiple case-study yielded a number of new insights concerning the efficacy of comprehensive CF. For instance, in line with suggestions in the literature (e.g. Ferris, 2004; Loewen, 2004), a learner’s level of successful CF uptake showed to be predictive of a pupil’s success in acquiring a target form. Two other interesting findings pertain to the correctability of different types of errors. Firstly, all error types proved to be amenable to CF, as long as errors were not fossilized. Secondly, learners only benefited from indirect correction when the feedback was targeting errors which could be solved by applying a clear-cut rule (i.e. the omission of an article in front of a Dutch noun in a definite context). Direct CF proved to be more effective in tackling learners’ idiosyncratic errors (e.g. the use of a wrong article in Dutch).
Chapter 6: Conclusion
The final chapter synthesizes the results of the three empirical studies (Chapters 3, 4 and 5), sketches out their theoretical and pedagogical implications, and discusses some directions for further research. The most important findings of the present studies can be summarized as follows: (i) Unfocused CF enables L2 learners to improve the accuracy of their written output; the research presented in this thesis showed that learners were not only able to make effective use of CF during revision, but that comprehensive CF also resulted in L2 learning. (ii) All types of errors are amenable to CF. (iii) The relative efficacy of direct and indirect CF is dependent on the type of error that is targeted. (iv) CF does not come with any side-effects that harm L2 learners’ accuracy development. (v) In-depth analyses of CF effectiveness can be considered a valuable addition to the more common global measures of accuracy development.
Samenvatting

Dit proefschrift onderzoekt de effecten van correctieve feedback op de schrijfvaardigheid van anderstalige leerlingen in het voortgezet onderwijs.

Hoofdstuk 1: Introductie

Hoofdstuk 1 schetst de context waarbinnen het gepresenteerde onderzoek is uitgevoerd en introduceert de concepten die in dit proefschrift een centrale rol spelen.


Eén van de cruciale vragen is nu hoe die feedback op het taalgebruik van tweedetaalleerders eruit moet zien. Een gangbare vorm van talige feedback is correctieve feedback (CF). CF kan gedefinieerd worden als elke indicatie aan de leerder dat zijn gebruik van de doeltaal niet correct is (Lightbown & Spada, 2006, p. 197). Hoewel CF in de onderwijspraktijk veelvuldig gebruikt wordt, is de effectiviteit van deze feedbackmethode omstreden. Sommige onderzoekers verwachten zelfs dat CF eerder een schadelijk dan een gunstig effect heeft op het proces van tweedetaalverwerving (bijv. Truscott, 1996). De vragen die in dit proefschrift centraal staan, zijn of, hoe en wanneer schriftelijke CF een positief effect heeft op de taalvaardigheid van tweedetaalleerders.
Schriftelijke CF kan veel verschillende vormen aannemen (zie tabel 1.1 op p. 4). De feedbacktypen die in dit boek onderzocht worden, zijn een directe en een indirecte variant van niet-gefocuste CF. Met niet-gefocuste CF wordt bedoeld dat alle fouten in een tekst van feedback voorzien worden. Deze feedbackmethode staat tegenover een gefocuste benadering, waarin een docent ervoor kiest slechts een (beperkt aantal) type(n) fouten te corrigeren (bijv. alleen lidwoordfouten). Het onderscheid tussen directe en indirecte CF heeft betrekking op de rol van de leerder in het correctieproces. Geeft een docent directe CF, dan identificeert hij zowel de fout als de doelvorm (zie voorbeeld 1). In indirecte correctie daarentegen, wordt de fout geïdentificeerd, maar is het aan de leerder de juiste doelvorm af te leiden. Indirecte CF kan op verschillende manieren gegeven worden; fouten kunnen bijvoorbeeld onderstreept of geturfd worden. In het huidige onderzoek werden fouten voorzien van verschillende foutcodes (zie voorbeeld 2 en Appendix C).

### Voorbeeld 1: directe correctieve feedback

| Je moet het trui niet chemisch reinigen. | Je moet het trui niet chemisch(S) reinigen. |
| de chemisch | ( _ = verkeerd woord, S = spelfout) |

De studies die in Hoofdstukken 3, 4 en 5 worden beschreven, onderzoeken de effectiviteit van directe en indirecte niet-gefocuste CF door vast te stellen welk effect schriftelijke feedback sorteert op de linguïstische correctheid (accuracy) van teksten. Met linguïstische correctheid wordt bedoeld dat een tekst op grammaticaal, lexicaal en orthografisch gebied voldoet aan de eisen die de doeltaal stelt. Meer concreet probeert deze dissertatie een antwoord te geven op de volgende zes vragen:

1. Leidt niet-gefocuste CF tot linguïstisch correcter taalgebruik in een gereviseerde tekst?
2. Leidt niet-gefocuste CF tot linguïstisch correcter taalgebruik in nieuwe teksten?
3. Welk correctiemethode is het meest effectief: directe of indirecte CF?
4. Welke factoren beïnvloeden de effectiviteit van niet-gefocuste CF?
5. Heeft niet-gefocuste CF schadelijke neveneffecten?
6. Hoe en wanneer profiteren individuele leerders van niet-gefocuste CF?
Hoofdstuk 2: Theoretisch en empirisch kader

Hoofdstuk 2 schetst het theoretische en empirische kader van de studies in dit proefschrift en doet dat vanuit een cognitief perspectief op tweedetaalverwerving. Het eerste deel van het hoofdstuk legt uit waarom vanuit een theoretisch oogpunt te verwachten valt dat CF bijdraagt aan de ontwikkeling van een tweede taal. Daarna wordt uiteengezet welke bezwaren tegenstanders zien in het gebruik van CF. Vervolgens bespreekt het hoofdstuk een aantal controverses met betrekking tot de relative effectiviteit van verschillende CF methoden en de corrigeerbaarheid van verschillende typen fouten. Tot slot volgen uit een bespreking van eerder onderzoek de kwesties die in dit proefschrift onderzocht worden.


Behalve over de vraag of CF effectief is, zijn de meningen ook verdeeld over de vragen welke feedbackmethode het meest effect heeft en of alle typen fouten in dezelfde mate ontvankelijk zijn voor CF. Zo worden in de literatuur uiteenlopende voorspellingen gedaan met betrekking tot de relatieve effectiviteit van gefocuste vs. niet-gefocuste CF en directe vs. indirecte correctie en stelt Truscott (1996) dat niet-grammaticale fouten (bijv.
spelfouten) eventueel te corrigeren zijn, maar grammaticale fouten (bijv. fouten in zinsconstructie) zeker niet.

Wat eerder onderzoek al aantoonde, is dat CF, in welke vorm dan ook, leerders helpt de talige correctheid van een tekst te verbeteren tijdens revisie. De belangrijkere en interessantere vraag is echter of correctie ook een leereffect bewerkstelligt: is het zo dat de feedback die een leerder ontvangt op één bepaalde tekst nog effect heeft op de correctheid van nieuw te schrijven teksten? Recente studies lieten zien dat gefocuste CF inderdaad een langetermijn-effect oplevert. Onderzoek naar eventuele leereffecten van niet-gefocuste CF daarentegen, kampte met methodologische tekortkomingen en kwam tot tegenstrijdige conclusies. Dat laatste geldt ook voor studies die de effectiviteit van directe en indirecte CF vergeleken. Kwesties als de potentiële schadelijke neveneffecten van CF of factoren die de effectiviteit van correctie mogelijk beïnvloeden (bijv. het type fout dat gecorrigeerd wordt, het opleidingsniveau van een leerder, het onderwerp van een schrijftaak), zijn bovendien nog te weinig (systematisch) onderzocht. De drie empirische studies in dit proefschrift richtten zich op het beantwoorden van deze open vragen.

**Hoofdstuk 3: Kleinschalige studie**

Hoofdstuk 3 beschrijft een kleinschalige studie (N = 66) naar het effect van directe en indirecte niet-gefocuste feedback.

Het experiment bestond uit drie sessies. In de eerste sessie (week 1 - pretest) schreven alle leerlingen twee teksten over twee verschillende onderwerpen (zie Appendix A). In de tweede sessie (week 2 - interventiesessie) reviseerden de leerlingen hun teksten op basis van directe dan wel indirecte niet-gefocuste CF, verbeterden hun teksten zonder beschikbare feedback of oefenden hun schrijfvaardigheid door twee nieuwe teksten te schrijven. In de derde sessie (week 3 - posttest) schreven alle leerlingen – ongeacht hun onderzoeksgroep – nogmaals twee teksten. Om de effectiviteit van de verschillende condities te kunnen bepalen, werd voor alle teksten de mate van linguïstische correctheid (accuracy) berekend op basis van foutratio’s (nl. (aantal linguïstische fouten/totaal aantal woorden) x 10).

De resultaten van deze eerste studie lieten zien dat niet-gefocuste CF tweedetaal leerders helpt de talige correctheid van hun geschreven output te vergroten. Zowel leerlingen wiens fouten direct gecorrigeerd werden als leerlingen die indirecte CF ontvingen, produceerden tijdens de interventiesessie teksten van hogere kwaliteit dan leerlingen in de twee controle condities. Directe CF leverde bovendien een leereffect op: de posttest teksten van leerlingen die directe feedback ontvingen, bevatten significant minder fouten dan die van leerders in de oefengroep en zelfcorrectiegroep. Deze bevindingen bieden dan ook geen ondersteuning voor Truscott's (1996) stelling dat extra taalproductie (of schrijfoefening) een gunstiger effect heeft op tweedetaalverwerving dan correctie en revisie. Ten slotte bleek het onderwerp van een schrijftaak niet van invloed te zijn op de manier waarop of mate waarin leerlingen profiteerden van CF.

Hoofdstuk 4: Grootschalige studie
In Hoofdstuk 4 worden de resultaten van een grootschalige studie (N = 268) naar de effectiviteit van niet-gefocuste schriftelijke CF gepresenteerd.

Vier middelbare scholen met een meertalige leerlingenpopulatie in de Randstad deden mee aan het onderzoek. De opzet van deze studie kwam grotendeels overeen met de methodologie van het kleinschaligere onderzoek zoals beschreven in Hoofdstuk 3. Op een vijftal punten werden wijzigingen in de onderzoeksopzet doorgevoerd: (i) Omdat uit de eerste studie bleek dat het onderwerp van een schrijftaak geen invloed heeft op de effectiviteit van CF, werd deze variabele niet meegenomen in de grootschalige studie. (ii) Om te kunnen bepalen of het opleidingsniveau van een leerder van invloed is op de effectiviteit van CF, namen niet alleen vmbo-tl leerlingen maar ook havo leerlingen deel aan het tweede onderzoek. (iii) De grootschalige studie omvatte vier in plaats van drie
meetmomenten. Vier weken na de interventiesessie werd een uitgestelde posttest afgenomen om meer inzicht te vergaren in de duurzaamheid van CF effecten. (iv) De studie die in Hoofdstuk 4 wordt beschreven, onderzocht niet alleen het effect van CF op de algehele talige correctheid van teksten. In de analyses werd ook onderscheid gemaakt tussen grammaticale en niet-grammaticale correctheid om Truscotts (2001; 2007) hypothese te kunnen toetsen dat grammaticale fouten niet corrigeerbaar zijn. (v) Ten slotte werd behalve naar maten van linguïstische correctheid ook gekeken naar de lexicale diversiteit en structurele complexiteit van teksten, om vast te kunnen stellen of CF leidt tot vermijding en simplificering van het taalgebruik.

Uit de resultaten van de grootschalige studie bleek wederom dat niet-gefocuste CF een positief effect heeft op de correctheid van zowel gereviseerde als nieuwe teksten. Bovendien bleek het effect van CF duurzaam te zijn; leerlingen wiens fouten gecorrigeerd werden, presteerden ook een maand na ontvangst van de feedback nog beter dan leerders in de controlegroepen. Een belangrijke bevinding was verder dat zowel grammaticale als niet-grammaticale fouten te corrigeren zijn, zij het met behulp van verschillende feedbackmethoden. Resultaten lieten zien dat alleen directe correctie leidt tot verbetering op grammaticaal niveau en dat het effect van CF op niet-grammaticale fouten duurzamer is als leerders indirecte CF ontvingen. Het opleidingsniveau van leerlingen bleek geen invloed te hebben op de effectiviteit van CF. Tot slot werd geen enkele aanwijzing gevonden voor de potentiële schadelijke effecten die door tegenstanders aan CF toegeschreven worden. Ten eerste bleek correctie leerders er niet toe aan te zetten meer complexe structuren of woordkeuze te vermijden. Ten tweede presteerden leerlingen die CF ontvingen in alle fases van het onderzoek beter dan leerders in de oefengroep. Het is dus niet het geval dat CF als een nutteloze activiteit beschouwd moet worden die kostbare tijd en energie opslokt die beter aan het schrijven van nieuwe teksten besteed had kunnen worden.

**Hoofdstuk 5: Kwalitatieve studie**

In Hoofdstuk 5 wordt verslag gedaan van een meervoudige casestudie (N = 4) naar het effect van directe en indirecte niet-gefocuste CF.

De studies in Hoofdstukken 3 en 4 leveren een belangrijke bijdrage aan het debat rondom de effectiviteit van CF door te laten zien dat niet-gefocuste CF een positief effect heeft op het proces van tweedetaalverwerving. Toch laten kwantitatieve studies als deze een aantal interessante kwesties onbeantwoord. Ten eerste biedt kwantitatief onderzoek
SAMENVATTING
geen inzicht in *hoe* en *wanneer* individuele leerders profiteren van CF (Storch & Wigglesworth, 2010). Ten tweede onthulde de studie in Hoofdstuk 4 dat zowel niet-grammaticale als grammaticale fouten corrigeerbaar zijn, maar blijft onduidelijk hoe afzonderlijke fouttypen binnen deze brede categorieën op CF reageren. Ten slotte suggereert Bruton (2009a, 2010) dat globale maten van linguïstische correctheid (zoals de foutratio’s in Hoofdstukken 3 en 4) wellicht geen adequaat en compleet beeld geven van de effecten die CF teweegbrengt. Eén van de argumenten die hij voor deze stelling aandraagt, is dat globale maten een ‘alles-of-niets’ criterium (nl. correct/incorrect) toepassen. Als CF tot gedeeltelijke vooruitgang leidt, zal dit positieve effect daarom niet in de globale maat geregistreerd worden.

Om bovengenoemde redenen wordt de effectiviteit van directe en indirecte niet-gefocuste CF in Hoofdstuk 5 op kwalitatieve wijze geanalyseerd. Voor deze meervoudige casestudie werden vier leerlingen geselecteerd uit de deelnemersgroep van het in Hoofdstuk 4 beschreven grootschalige onderzoek: twee leerlingen uit de groep die directe CF ontving en twee leerders wiens fouten indirect gecorrigeerd werden. Om inzicht te krijgen in hoe en wanneer leerders al dan niet profiteren van correctie, werden leerders gekozen die op verschillende manieren reageerden op CF. Voor beide feedbacktypen werd één leerling gekozen van wie de teksten minder fouten bevatten na CF dan ervoor en één leerling die een toename liet zien in het aantal fouten tussen de pretest en de posttesten.

Voor iedere leerling werden zijn vier schrijfproducten (nl. pretest tekst, gereviseerde versie van pretest tekst, posttest tekst en uitgestelde posttest tekst) aan een gedetailleerde vergelijking onderworpen. Wanneer een feedbackepisode tot correcte revisie of herformulering leidde, werd dit als evidentie van *feedbackuptake* geïnterpreteerd. Het correcte gebruik van een gecorrigeerde vorm in een nieuwe tekst werd als bewijs beschouwd van *feedbackretentie* of *taalontwikkeling*. Fouten in de posttest teksten die geen enkel verband hielden met de fouten die een leerling in zijn pretest tekst had gemaakt, werden als *nieuwe fouten* aangetekend. Om inzicht te kunnen krijgen in de corrigeerbaarheid van afzonderlijke fouttypen, werden alle fouten geclassificeerd in vier domeinen (morfosyntactische, lexicale, orthografische en pragmatische fouten), elk verder uitgesplitst in een aantal subtypen (zie tabel 5.1 op p. 98).

De kwalitatieve benadering in Hoofdstuk 5 leverde zowel andere als aanvullende inzichten op in vergelijking tot de kwantitatieve studies in Hoofdstukken 3 en 4. Ten eerste bleek dat het positieve effect van CF wordt onderschat wanneer alleen gekeken wordt naar globale maten van linguïstische correctheid. De vooruitgang die leerders boekten, bleek

Bovendien volgden uit de meervoudige casestudie een aantal nieuwe inzichten. Zo bleek succesvolle feedbackuptake, in overeenstemming met suggesties in de literatuur (bijv. Ferris, 2004; Loewen, 2004), een goede voorspeller te zijn voor feedbackretentie. Twee andere interessante bevindingen betroffen de corrigeerbaarheid van verschillende typen fouten. Ten eerste bleken alle typen fouten ontvankelijk te zijn voor CF, zolang fouten niet gefossiliseerd waren. Ten tweede profiteerden leerders alleen van indirecte CF als de feedback gericht was op fouten die op basis van een duidelijke regel te verbeteren waren (bijv. het ontbreken van een lidwoord voor een zelfstandig naamwoord in een definiete context). In het bestrijden van meer idiosyncrasische fouten (bijv. het gebruik van een verkeerd lidwoord) bleek directe CF effectiever.

**Hoofdstuk 6: Conclusie**

In het laatste hoofdstuk worden de resultaten van de drie empirische studies (Hoofdstukken 3, 4 en 5) samengebracht, de theoretische en pedagogische implicaties uiteengezet en een aantal suggesties besproken voor verder onderzoek. De belangrijkste conclusies die uit de onderhavige studies naar voren komen, laten zich als volgt samenvatten: (i) Niet-gefocusste CF helpt tweedetaalleerders de linguïstische correctheid van hun schriftelijke taalgebruik te vergroten; de studies toonden niet alleen aan dat leerders feedback kunnen gebruiken tijdens revisie, maar ook dat niet-gefocusste CF een leereffect oplevert. (ii) Alle typen fouten zijn corrigeerbaar. (iii) De relatieve effectiviteit van directe en indirecte CF is afhankelijk van het type fouten waarop de feedback gericht is. (iv) CF leidt niet tot schadelijke neveneffecten. (v) Gedetailleerde analyses van CF effectiviteit vormen een waardevolle aanvulling op de gebruikelijke globale maten.
Catherine van Beuningen was born in The Hague on December 4th 1979. After obtaining her gymnasium diploma from the Christelijk College ‘De Populier’ in The Hague in 1998, she studied General Linguistics at Utrecht University. Specializing in syntax and psycholinguistics, she wrote an MA-thesis on the cognitive processing of Dutch and Frisian pronouns. After graduation in 2004, she taught Dutch language, culture, and linguistics at the Károli Gáspár University in Budapest (Hungary), and Dutch as a second language to expats and immigrants at Direct Dutch (The Hague) and ROC Mondriaan (Delft). Working with second language learners sparked her interest in applied linguistics. In 2006, Catherine van Beuningen started her Ph.D. project at the Amsterdam Center for Language and Communication (ACLC), which resulted in the present dissertation. From 2007 onwards, she also taught courses at the Amsterdam University of Applied Sciences (HvA) and in the Dutch Linguistics department of the University of Amsterdam. She currently teaches different BA and MA modules at the University of Amsterdam and at the Amsterdam University College. She is also a board member of the Dutch Association for Applied Linguistics (Anéla).