Writing processes, text quality, and task effects

Empirical studies in first and second language writing

Published by
LOT
Janskerkhof 13
3512 BL Utrecht
The Netherlands

phone: +31 30 253 6006
fax: +31 30 253 6406
e-mail: lot@let.uu.nl
http://www.lotschool.nl

Cover illustration by Wilma van Weijen – van Iersel
ISBN 978-90-78328-75-9
NUR 616

Writing processes, text quality, and task effects

Empirical studies in first and second language writing

Schrijfprocessen, tekstkwaliteit, en taakeffecten

Empirisch onderzoek naar schrijven in moedertaal en tweede taal

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor
aan de Universiteit Utrecht
op gezag van de rector magnificus, prof. dr. J.C. Stoof,
ingevolge het besluit van het college voor promoties
in het openbaar te verdedigen
op vrijdag 30 januari 2009 des ochtends te 10.30 uur

door

Daphne van Weijen

geboren op 27 maart 1978 te Soest

Promotoren: Prof. dr. H. van den Bergh
Prof. dr. G.C.W. Rijlaarsdam
Prof. dr. T.J.M. Sanders

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

"Oh what things we writers come up with in order to avoid doing what we love to do most: write." [1]
Maria Goos

Although carrying out this research project and writing this thesis was tough at times, I enjoyed it immensely and wouldn't have missed it for the world! But I would not have been able to complete it or would not have found it so enjoyable without the help and support of many people along the way. I thank you all! However, in the hope that I don't forget anyone, a number of people deserve a more extended thank you.

First of all, this thesis was written in the *plural we*, with good reason. Without the guidance provided by three wise men, my supervisors: Huub van den Bergh, Gert Rijlaarsdam, and Ted Sanders, this thesis would not have developed into the book that lies before you. They each contributed their own expertise to this project and I am truly grateful for all that I have learned from them over the past five years. Huub, thank you for being my daily supervisor and research mentor. Your seemingly endless optimism regarding my project and most of the tasks on my endless to-do lists (*"Appeltje eitje!"*), and your tireless explanations related to research in general and methodology in particular, helped me more than you will ever know. Thanks to your sound advice I never ended up with egg on my face, even though I often feared the worst. Gert, thank you for being the best long-distance supervisor an AiO could hope for. Even though we rarely discussed my project face-to-face, your supervision via e-mail was crucial for its success. Thank you for your detailed feedback and insightful comments on all my papers and for introducing me to the EARLI SIG Writing community. Ted, thank you for your text-linguistic input, your advice, and for letting my research project '*migrate*' in other directions than you might have initially expected.

Second, I wish to thank Martin Everaert, Maaike Schoorlemmer and UiL-OTS for giving me the opportunity to carry out my research in Utrecht, for providing the funding for research assistance during my main experiment, and for making it possible for me to attend so many inspiring conferences. Furthermore, I wish to thank Martine Braaksma, Simon Cook, Rias van den Doel, Mark de Jong, Marleen Kieft, Mariëlle Leijten, Iris Mulders, Marion Tillema, Theo Veenker, and the teachers and students of the pre-test at the Hogeschool van Amsterdam for their help and input at various stages of the main experiment. I also wish to thank my research assistants, Ellen Bruens and Mirjam Slieker, as well as William van Heeteren, for their help in transcribing the think-aloud protocols, and Luuk van Waes and Mariëlle Leijten for developing Inputlog, which made the enormous transcription task, and my life, a whole lot easier. I would also like to thank all the BA English students who participated in the pre-test and the main experiment, and all the BA and MA students who helped code the protocols and judge the quality of the essays.

---

[1] *"Wat wij, schrijvers, allemaal verzinnen om maar niet dat te doen wat we het liefste doen: schrijven."* Maria Goos, Volkskrant Magazine, pg. 15, March 25th 2006

Third, Jacqueline Evers-Vermeul, Rosie van Veen, Loes Ruiter-Bloemendal, and Merel van Weijen, were kind enough to read sections of this thesis or provide linguistic advice, for which I am very grateful. Any remaining errors are of course my own. Furthermore, I would like to thank Catherine van Beuningen for helping me obtain a copy of an article which proved essential for the literature search related to chapter 6.

Although I often worked on my research in relative isolation, especially while transcribing protocols, I never felt alone, thanks to my colleagues. Therefore, I wish to thank my roommates (in chronological order): Marion Tillema, Judith Kamalski, Jentine Land, Sanne Elling, Rosie van Veen, Judith Evertse, and Elena Tribushinina, and fellow AiO's: Naomi Kamoen, Nina Versteeg, and Judith van Wijk, for being such good company, for their endless support, and for all the laughs, lunches, pots of tea, and sweets we shared. Furthermore, I wish to thank all my colleagues of the afdeling Taalbeheersing and UiL-OTS for making me feel so welcome and for taking the time to discuss my work with me during conferences, lunch breaks, and informal chats.

Besides my colleagues in Utrecht, I would also like to thank Eva Lindgren, Mariëlle Leijten, Carien Bakker, Catherine van Beuningen, and the extended members of the ILO/GSTL Language and Literature Research team including: Martine Braaksma, Hein Broekkamp, Michel Couzijn, Tanja Janssen, Marleen Kieft, Talita Groenendijk, Mariet Raedts, Elke van Steendam, and Anne Toorenaar, for our interesting research-related discussions and for their company at conferences.

Going back in time, I would also like to take this opportunity to thank Huub van den Bergh, Wolfgang Herrlitz, Jacomine Nortier, Thijs Pollmann and Peter Slagter, for sparking my interest in sociolinguistics, language teaching, and doing research. If it weren't for their inspiring lectures and the advice they gave me as a student, I might never have started this project in the first place.

Finally, a great big THANK YOU goes to my family and all my friends for their support and for their patience, especially in the final year, when I was often absent (physically, mentally, or both). In addition, I would like to thank Sanne Elling and Merel van Weijen for agreeing to be my paranimfs. Sanne, thank you for sharing the ups-and-downs of AiO-life with me over the past few years and for your friendship. Our lunchtime walks and talks often helped me see things in a new light and gave me much needed energy at times when I was feeling overwhelmed. Merel, you're the best sister and friend anyone could ask for. It means a lot to me to have you by my side during my defence. Pappa en Mamma, I want to thank you for all the opportunities and support you've given me over the years, and for always believing that I could do whatever I set my mind to. I'm finally starting to believe you might be right. My last thank you and all my love goes to Sam. Lief, bedankt voor alles….

Daphne

Utrecht, November 17th 2008

CHAPTER 1

INTRODUCTION

Writing a text usually entails a lot of hard work. Indeed: *"... generally, writing anything but the most routine and brief pieces is the mental equivalent of digging ditches"* (Kellogg, 1994, p. 17). Therefore, it seems almost miraculous that writers manage to produce any texts at all, let alone texts of any genuine quality. Since the early 1980's, writing research has focused on trying to solve this puzzle. How do writers write their texts, and more importantly, how are good texts produced? To date, most studies which explored and attempted to unravel this problem focussed on the cognitive processes involved. The aim of the present series of studies, on the other hand, is to advance our understanding of the writing process by *relating* process and product characteristics to each other. In other words, how does the way in which writers write influence the quality of the texts they produce? The theoretical starting point of this study is the Descriptive interactive process model presented in Figure 1 (Van den Bergh, Rijlaarsdam, Janssen, Braaksma, Van Weijen, & Tillema, in press, p. 5; see also Rijlaarsdam, et al., 2008).

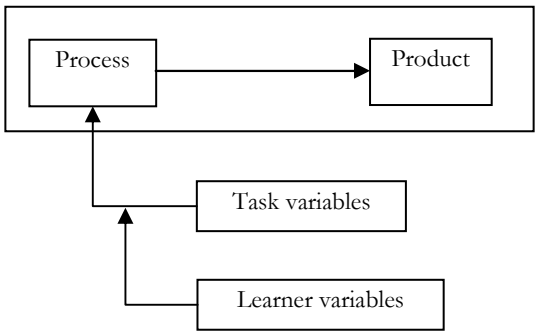

Figure 1: The Descriptive interactive process model (for reading & writing) (Reproduced from: Van den Bergh, et al., in press, p. 5).

The model in Figure 1 shows that the relationship between the writing process and its product, the resulting text, is influenced on the process-side by – at least – two types of variables: task variables and learner variables. In our studies, the focus will be on one particular task variable, the language in which the writing process is carried out: a writer's first (L1) or his second language (L2). This variable was chosen, because research has shown that writers often produce texts of lower

quality when they write in L2 than might be expected based on their L1 writing expertise (see for example Silva, 1993; Couzijn, Rijlaarsdam, & Van den Bergh, 2002; Van Weijen, Van den Bergh, Rijlaarsdam, & Sanders, 2005a). This is a growing problem, because being able to write well in L2 is becoming more and more important in Higher Education. English is increasingly becoming the medium of instruction, particularly in Master Programmes at Universities across Europe. Therefore, being able to write well in English, is becoming a skill that is essential for academic success. In a study on the relationship between cognitive activities and text quality during writing in L1 and L2, Couzijn, et al. (2002) found that some writers who produced high quality texts in L1, produced texts of much poorer quality in L2, while other writers produced relative high quality texts in both languages. To illustrate this finding, some excerpts from their data are shown in Table 1.

Table 1: Examples of writing in L1 and L2 from data collected by Couzijn, et al. (2002)

	L1 (Dutch)	*FL (English)*
Writer A	*"… Er is dus zeker wel stress onder scholieren (eigenlijk heeft iedereen wel eens last van stress), maar dat kan je het hoofd bieden door je aan bovenstaande punten te houden." [Translation: "… Pupils certainly suffer from stress (everyone suffers from stress now and again), but you can cope with it by sticking to the aforementioned points."]*	*"…I believe that – first of all – you have to be married, if you want to have kids, because the child has a right to a legal father and mother. And if you are married, I think you have to think a lot before having a child or not, because it will affect you for the rest of your life."*
Writer B	*"… Het lijkt mij gezelliger om samen te wonen, want dan is er iemand die voor je zorgt en je kunt samen leuke dingen gaan doen, dat is veel gezelliger." [Translation: "… I think it would be nicer to live together, because then there's always someone who takes care of you and you can do enjoyable things together, that's much more fun."]*	*"… A word like stress means, just like success, is (I thik so) rite, becaus you will succes when you press to stress. than go you beter you best, and that is good."*

The writers in Couzijn, et al.'s study were high school students, approximately 15 years old, with very limited experience with writing in English. At face value, Writer

A is obviously a much better writer in FL than Writer B, although it is hard to determine based on such brief excerpts. But Writer B's text is clearly much lower in quality, for example, due to the fact that her FL text appears to be a word-for-word translation from L1, which makes the text very difficult for readers to comprehend. In addition, the presence of spelling and punctuation mistakes also make the intended meaning of the sentence in the excerpt very unclear, although it appears to be that stress is not a bad thing, because writers often work better under pressure. So, although Writer B appears to write good quality texts in L1, her FL writing is very hard to read and clearly much lower in quality. But why did she produce texts of such poor quality in English? One possible explanation is that the decrease in text quality might be related to differences in the orchestration of the writing process in L1 and L2, because research suggests that the hierarchical structure of the text, or the lack of structure in Writer B's case, *"reflects the writer's cognitive activities during writing"* (Sanders, Janssen, Van der Pool, Schilperoord, & Van Wijk, 1996, p. 473; see also Sanders & Schilperoord, 2006; Sanders & Van Wijk, 1996). Therefore, in this study, we will compare how writers produce their texts in L1 and L2, in order to determine how task variables such as language (L1 vs. L2) and learner variables such as L1 writing skill and L2 proficiency influence the orchestration of the writing process, and thus influence the quality of the texts produced.

The following sections provide a brief overview of several writing process models, followed by more details on the rationale for this study, and an overview of the themes that will be addressed.

Models of the writing process

The writing process clearly plays a central role in Van den Bergh, et al.'s (in press) Descriptive interactive process model (see Figure 1). But the model does not specify which components the writing process consists of and how these interact during text production. Researchers have attempted to provide a realistic and theoretically sound model of the writing process, since the early 1980's. The first and probably most well-known attempt at unravelling how the writing process unfolds is Flower and Hayes' model of the writing process (Flower & Hayes, 1980a; Hayes & Flower, 1980). This model contained three core components: the task environment, the writer's long term memory and the writing process. They provide some information on the task environment and long term memory components and how these are related to the writing process, but only in rather brief terms (see Hayes & Flower, 1980, p. 10 – 12). The focus is, not surprisingly, on the writing process component of the model, containing three main cognitive components: planning, translating, and reviewing, and a monitor component (Hayes & Flower, 1980, p. 12). The planning and reviewing components each contain several sub-processes, some of which, such as generating and organizing

are further specified in separate models (Hayes & Flower, 1980, p. 13 – 15). Finally, the monitor component is responsible for the orchestration of the execution of all these different activities (Flower & Hayes, 1980a, p. 39; Hayes & Flower, 1980, p. 19).

Over the years, several other models of the writing process have been presented, each with its own specific focus. Bereiter and Scardamalia (1987), for example, proposed a model that focused on writing development, from knowledge-telling to knowledge transforming (Bereiter & Scardamalia, 1987). Subsequently, two models were presented that introduced working memory as a component in the writing process (Hayes, 1996; Kellogg, 1996). Hayes (1996) presented a revised version of the original model (Hayes & Flower, 1980), which included two main components instead of the original three: the individual and the task environment. The writer's long-term memory and the writing process, which were originally two separate components, were combined to form a larger category, the individual, which also included a motivation component and the writer's working memory (Hayes, 1996). Although this revised version of the model was structured rather differently, it still resembled the old model to a large extent. In Kellogg's model, the focus was less on the writer as an individual, but more on the way in which the activities that occur during writing are related to specific subcomponents of working memory: the central executive, the visuo-spatial sketchpad and the phonological loop (Kellogg, 1996). As a result, it provided a valuable basis for more experimental research aimed at determining how working memory can facilitate and/or impede the execution of the writing process (see for example, Kellogg, 1988; Kellogg, Olive, & Piolat, 2007; Levy & Ransdell, 1995; Pélissier & Piolat, 1998; Piolat, 1999; Ransdell, Levy, & Kellogg, 2002).

Differences between these four models notwithstanding, they are all complementary to some extent (see Alamargot & Chanquoy, 2001, p. 21). All the models include a number of cognitive components, related to three main cognitive activities: planning, formulating and revising. In addition, they all contain a control mechanism which orchestrates the overall execution of the writing process and, finally, they all stress the potentially inhibiting role of working memory (see Alamargot & Chanquoy, 2001; O'Brien, 2004, p. 3).

The present study is partly based on the common elements of all of these models, but was primarily carried out as a result of questions regarding writing process research raised by Rijlaarsdam and Van den Bergh (1996). They also proposed a *"probabilistic model of writing processes"* (Rijlaarsdam & Van den Bergh, 1996, p. 108), which contained an executive component and a management component. The most distinctive aspect of their model is the focus on changing task representations as a key feature of the writing process. As the writing process develops, writers need to adapt to changes in the task situation, in order to carry out the writing task successfully. *"Because of the impossibility to observe mental*

representations of task situations, we propose to indicate changing task representations by the variable time" (Rijlaarsdam & Van den Bergh, 1996, p. 107). In studies that followed, they consistently stressed the importance of time as a key factor in the writing process (see for example, Breetvelt, Van den Bergh, & Rijlaarsdam, 1994; Rijlaarsdam & Van den Bergh, 2006; Van den Bergh & Rijlaarsdam, 2001; Van den Bergh, et al., in press). In their 1996 study, Rijlaarsdam and Van den Bergh raised many pertinent questions, some of which are the focus of the present study, as will be described below.

Defining process-product relations

As mentioned earlier, writers' L2 texts often appear to be lower in text quality than was expected based on the quality of their L1 texts (see Couzijn, et al., 2002; Van Weijen, et al., 2005a). This could be due to the influence of language (L1 versus L2) on writers' orchestration of the writing process (see also Van den Bergh, Herrlitz, & Klein Gunnewiek, 1999). Therefore, the main focus of this research project is on the influence of the language writers write in on process-product relations, or more specifically on how what writers do when they write influences the quality of the texts they produce.

In the present study, the writing process is defined as a set of 11 cognitive activities which each contribute to the writing process as a whole in some way. This set of activities is largely based on the categories included in Flower and Hayes' (1980a) model, such as reading the assignment, planning, generating ideas, formulating, evaluating and revising (Hayes & Flower, 1980; see also Breetvelt, et al., 1994; Couzijn, et al., 2002; Rijlaarsdam, 1986; Van den Bergh, 1999; Van den Bergh & Rijlaarsdam, 1999; Van der Hoeven, 1996a). Previous research often equated experience with effectiveness, which means that experienced writers were considered to be good writers, while novice writers were categorized as weaker writers (see also Torrance, 1996). Although this line of reasoning does have some merit, it seems intuitively far more logical to determine the effectiveness of writing processes by relating them to more concrete output measures. Therefore, Rijlaarsdam and Van den Bergh proposed that future research should use: *"the resulting text quality of the writing process under study as a proxy-variable or indicator of effectiveness"* (Rijlaarsdam & Van den Bergh, 1996, p. 115). In line with this idea, several other studies also suggested that it is important to include text quality *"as an indicator of the effectiveness of the processes under study"* (Roca de Larios, Marín, & Murphy, 2001, p. 528; see also Arndt, 1987, p. 257). Furthermore, research has also suggested that the reverse also applies. Text analysis can help shed light on the way in which text production takes place, because: *"cognitive writing activities are – at least to a certain extent – reflected in the hierarchical nature of the text produced"* (Sanders & Schilperoord, 2006, p. 400; see also Sanders, et al., 1996; Sanders & Van Wijk, 1996). Although we agree that a combination of text analytical and on-line

measures of writing seems fruitful (see Sanders, et al., 1996, p. 491), we will focus mainly on on-line measures of writing processes in the present study. For the purpose of this study, the product-side of the equation is defined as the quality of the texts that are produced, which will be assessed in a rather conventional way by a panel of raters. However, in Chapter 4 we will present an attempt to integrate text characteristics such as structure and argumentation in our assessment of text quality (see also Sanders & Schilperoord, 2006). Finally, process-product relations are studied in order to determine the effect of the way in which cognitive activities are carried out on the quality of the resulting text. In other words, we wish to determine how variations in the writing process affect the quality of the product (i.e., the text produced; see Figure 1).

THE RATIONALE FOR THE STUDY

Rijlaarsdam and Van den Bergh (1996) proposed an agenda for writing research, which included three different categories of questions, related to: descriptive research on writing processes, descriptive research on writing processes and text quality, and to explanatory research on effective writing processes, which included cognitive skills as a potential explanatory variable (Rijlaarsdam & Van den Bergh, 1996, p. 115 – 116). In this study, we will attempt to answer some of the questions they raised, in relation to several different themes that have emerged in writing research from 1996 onwards. In the following section, we will discuss the four main themes of this thesis, of which theme one, related to language as a task feature, is the most important. The other three themes are all connected to this main theme in different ways.

Theme 1: Writing in L1 and L2

The main focus of this study is to determine how task variables influence process-product relations through their influence on the orchestration of the writing process (see Figure 1). The task variable under scrutiny will be the language in which the writing process is carried out. Most of the research to date has focused on L1 writing or on L2 writing, but rarely on both. In addition, the few studies that did focus on both were often based on a single-task design (see for example, Arndt, 1987; Beare & Bourdages, 2007; Raimes, 1994; Roca de Larios, Murphy, & Marín, 2002; Silva, 1993) or on a between group comparison instead of on within writer comparisons (e.g., Raimes, 1994). Several other studies, such as Thorson (2000), suggested that comparing the processes and products of the same writers in both languages is *"a significantly more effective research design for analyzing L1 and L2 writing processes than comparing across groups"* (Thorson, 2000, p. 156; see also Roca de Larios,

et al., 2002, p. 31; Zimmermann, 2000, p. 79). Consequently, in the present study, the same group of participants was asked to write texts in their L1 and their L2. The participants in this study are all mother-tongue speakers of Dutch, who all study English at university level. Strictly speaking, English is more likely to be a foreign language (FL) for these students than a second language (L2). However, as the world 'shrinks' due to people's increased mobility and increased contact through the use of new media and the internet, the boundaries between second and foreign languages are becoming more unclear. As a result, it has become common practice to refer to both second and foreign language writing as L2 writing (see for example, Broekkamp & Van den Bergh, 1996, p. 170; Cohen & Brooks-Carson, 2001, p. 184; Cook, 1996, p. 7; Schoonen, Van Gelderen, De Glopper, Hulstijn, Simis, Snellings, & Stevenson, 2003, p. 197). That is why we will refer to our participants as second language writers, in line with the UNESCO definition of a second language: *"A language acquired by a person in addition to his mother tongue"* (Cook, 1996, p. 7). Thus, for the purpose of this study, we will refer to Dutch as our participants' L1 and English as their L2.

In the present study, we wish to expand our understanding of the way in which writing in L1 and L2 influences the orchestration of the writing process. Are the results of earlier L1 writing research applicable to L2 writing as well? What effect does writing in L2 have on the orchestration of cognitive activities? These questions will be examined in Chapters 3 and 4. In addition, we will also approach this topic from a slightly different angle. When writers produce texts in L2, they automatically have two languages at their disposal to write in: their L1 and their L2. Therefore, the question is whether they use their L1 during writing in L2 and what influence that might have on the way the writing process unfolds. This topic will be addressed in Chapter 6.

Theme 2: Time as a factor in writing process research

Previous research on writing processes often focused on the frequency with which cognitive activities occurred during writing, rather than on the moment at which they occurred during the writing process (see for example, Flower, Hayes, Carey, Schriver, & Stratman, 1986; Piolat, 1999; Stevenson, 2005; Zimmermann, 2000). As mentioned earlier, we are convinced that time should be taken into account as a reflection of the changing task representation that the writer has to form while writing (Rijlaarsdam & Van den Bergh, 1996; Van den Bergh & Rijlaarsdam, 1999, p. 103). If we restrict our analysis, by focusing solely on the frequency with which activities occur, then two writers who carry out roughly the same amount of generating or revising activities may appear to have rather similar behaviour, while in fact they could be doing very different things (Van den Bergh & Rijlaarsdam, 1999, p. 103). For instance, a writer with start-up problems, who revises his first paragraph over and over again, might revise just as frequently as a writer with

fluent text-production who only revises near the end of the writing process. But the effect of their revisions on text quality could well be very different. Indeed, research on L1 writing has shown that the moment at which an activity occurs, and not its frequency of occurrence, is related to the quality of the text produced (see for example Breetvelt, et al., 1994; Van den Bergh & Rijlaarsdam, 2001). Hence, including time in analyses of writing processes is essential, because: *"If the moment an activity is employed is left out of the analysis, hardly any relation can be found between cognitive activities on* [the]*one hand and text quality on the other. Therefore, the moment a cognitive activity takes place is crucial."* (Rijlaarsdam & Van den Bergh, 2006, p. 42; see also Roca de Larios, Manchón, & Murphy, 2006; Roca de Larios, Manchón, Murphy, & Marín, 2008; Roca de Larios, et al., 2001). That is why, in line with the questions raised by Rijlaarsdam and Van den Bergh (1996) concerning descriptive research on writing processes, this study will focus on determining when cognitive activities occur within the writing process, and how the moment at which they occur varies between writers and between tasks (within writers), in L1 and L2 writing (see Chapters 2, 3, and 4).

Theme 3: Variation between tasks

Another issue related to writing process research was raised by Meuffels and Van den Bergh (2005). They argued that a single-task design, which was often used in previous research (see for example, Breetvelt, Van den Bergh, & Rijlaarsdam, 1996; Roca de Larios, et al., 2006; Roca de Larios, et al., 2001; Van den Bergh & Rijlaarsdam, 2001), is problematic, for several reasons. One reason is the fact that research based on a single task per writer cannot determine to what extent variation in process execution or in process-product relations is related to variation between writers or to variation between tasks. Thus the variation between writers is overestimated, which means that the generalizability of results from studies with a single-task design is seriously limited. Hence, the question is whether the findings of earlier studies on process-product relations are task dependent, which means that they only occur in specific tasks, or whether they are writer dependent, and are thus similar within writers across multiple tasks. So the question we will focus on is whether we can generalize findings from earlier studies across tasks in L1 and L2 (see Chapters 2 & 3; see also Roca de Larios, et al., 2008, p. 44).

Theme 4: Interaction between cognitive activities

In their 1996 paper, Rijlaarsdam and Van den Bergh also raised questions regarding the interaction between cognitive activities during the writing process (see Rijlaarsdam & Van den Bergh, 1996, p. 113, Table 6.1a). They suggested that *"activities (or a series of activities) relate to each other"* (Rijlaarsdam & Van den Bergh, 1996, p. 121). Thus, in subsequent studies, they examined the influence of combinations of activities on text quality, such as rereading the text produced so far

and generating ideas (Breetvelt, et al., 1996), and generating ideas in combination with several other activities (Van den Bergh & Rijlaarsdam, 1999, 2007; see also Rijlaarsdam & Van den Bergh, 2006). Several other studies have also suggested that cognitive activities might influence each other in some way during the writing process (see for example, Chenoweth & Hayes, 2001, p. 85; Hayes & Gradwohl Nash, 1996, p. 54; Kellogg, 2001, p. 223; Roca de Larios, et al., 2008, p. 44; Stevenson, 2005, p. 185; Rijlaarsdam & Van den Bergh, 2006; Van den Bergh & Rijlaarsdam, 2007; Van Weijen, 2006). Therefore, in the present study we will try to build on previous research by enlarging the unit of analysis in writing process research, to further investigate the influence that activities such as planning might have on the orchestration of the writing process (see Chapter 5).

OUTLINE OF THIS THESIS

To study the themes presented above, this research project includes a number of features which distinguish our studies from those of earlier research. First of all, the participants in this study are first year university students, who are considered to be rather experienced L1 writers, instead of high-school writers (Breetvelt, et al., 1994; Van den Bergh & Rijlaarsdam, 2001), or primary school children (Van der Hoeven, 1996a; 1999). Second, to study the influence of language on the writing process, multiple texts must be written by each writer in each condition. Otherwise it is impossible to determine whether differences between L1 and L2 are due to the language being written in or due to other differences between the tasks. Therefore, participants in this study were required to write eight tasks each, four in L1 and four in L2, instead of only one or two per condition.

In the next few chapters[1], we will present results from different analyses carried out on the data. First of all, it is important to note that all of the analyses reported in this thesis are related to theme 2, which means that they all include an operationalization of the changing task representation (see Chapters 2 – 6). Second, if we wish to compare how language affects the orchestration of writing processes, then it is important to first determine how the writing process is usually orchestrated, over writers and tasks, in L1. A comparison of L1 and L2 results is only possible if we are able to generalize findings across tasks in L1. Therefore, in Chapter 2, we will present an analysis of the effect of variation in the orchestration of cognitive activities on text quality in L1 writing. This is directly related to theme 3, which focuses on between-task variation. The main question, then, is, to what

[1] Unfortunately, there is some overlap in the method sections of Chapters 2 to 6, because these have all been submitted for publication in different international journals. The advantage this offers the reader, however, is that each chapter can be read on its own.

extent the execution of activities varies between writers and between tasks, and how this affects the quality of the texts produced.
In Chapter 3, the focus is on L2 writing, and on variation in L2 writing in particular, which means that this chapter is also directly related to theme 3. The question to be answered is whether writers vary how they write between tasks under influence of the changing task situation when writing in L2 and, if so, whether variation in the way they write is related to variations in text quality. Again, we can only compare results between L1 and L2 writing once we have determined how the writing process is usually orchestrated in both languages. Then we can generalize findings across writers and tasks, and thus compare them.

Chapter 4 contains a comparison of L1 versus L2 writing, which examines whether the relationship between the writing process and text quality changes when writers produce texts in a second or foreign language instead of in their first language. Therefore, this chapter forms the heart of this thesis, because it is directly related to theme 1 and because it has the strongest link with our main aim: to determine how language affects process-product relations.

The study presented in Chapter 5 is an exploratory study, related to themes 1 and 4, which focuses on the interaction between cognitive activities during the writing process in L1 and L2. The analysis is based on the idea that cognitive activities do not occur and function in isolation during the writing process. More specifically, it examines the influence that activities such as planning have on the writing process as a whole. Therefore, an analysis is carried out based on composing episodes, a larger unit of analysis, than was previously used in other studies. The aims of the study are, first of all, to determine from a methodological viewpoint whether the writing process can be divided into composing episodes, and second, to determine whether activities which occur at the start of episodes are able to influence the quality of the text produced in L1 and L2.

In Chapter 6, the analysis focuses on how writers use their first language (L1) while writing in their L2. This study examines when L1 use occurs during writing in L2, and how L1 use is related to text quality. In addition, it also examines how L1 writing skill and L2 proficiency, two different learner characteristics, influence both L1 use and text quality. Therefore, the study presented in this chapter is also related to theme 1.

Finally, in Chapter 7, we will summarize the main findings from each study. Furthermore, we will discuss methodological issues that arose from our results, provide suggestions for further research, and propose some implications for education.

CHAPTER 2

VARIATION IN THE ORCHESTRATION OF COGNITIVE ACTIVITIES AND TEXT QUALITY IN L1 WRITING

Abstract [1]

This study examines the variation in the orchestration of cognitive activities and text quality both between and within writers, in order to answer two questions: *(1) Do writers vary the orchestration of each cognitive activity during the writing process between tasks? (2) Does the relation between the orchestration of activities and text quality vary between tasks?* Twenty students each wrote four short argumentative essays in their L1 (Dutch) under think-aloud conditions. The analysis focused on four cognitive activities: Reading the assignment, Planning, Generating ideas and Formulating. Results indicate that writers do indeed vary the orchestration of cognitive activities between tasks, although the extent to which they do so is different for each activity. Second, the relation between the orchestration of activities and text quality varies both over time and between tasks. Therefore, we conclude that variations in the execution of activities between tasks appear related to variations in text quality, at least for the activities reported in the present study.

Keywords: Formulating; Generating; L1 Writing; Planning; Process Variation; Process-Product relations; Reading; Text quality; Text Production

Process – product relations

Writing process research emerged in the early 1980's, with the introduction of Flower and Hayes' model of the writing process (Flower & Hayes, 1980a; Hayes and Flower, 1980). Their text production model consists of three components: the writer's long-term memory, the task environment, and the writing process itself. This last component contains a monitor and three main cognitive components involved in writing: planning, formulating, and revising, which each contain several cognitive activities (Hayes & Flower, 1980). They suggest that the monitor controls the orchestration of these cognitive activities and that it is therefore responsible for the large differences between writers in the way they write (Flower & Hayes, 1980a, 1980b). Hayes proposed a revised version of the model in 1996. This new version

---

[1] Van Weijen, D., Van den Bergh, H., Rijlaarsdam, G., & Sanders, T. (2008). *Variation in the orchestration of cognitive activities and text quality in L1 writing.* Manuscript submitted for publication.

of the model contained two main components instead of three, and also included working memory as a central component. However, it still contained the same cognitive activities as the preceding model; although these were now grouped under different headings (see Hayes, 1996).

Initially, research on writing that attempted to build on the Flower and Hayes model focused on relating elements of the model to writer characteristics such as writing skill (novice vs. expert writers) or defined writing skill as high or low scores on a writing skill test (see for example, Breetvelt, 1988; Breetvelt, Van den Bergh, & Rijlaarsdam, 1994; Flower & Hayes, 1980b; Graham & Harris, 2000). In addition, research often focused on the role of a single cognitive activity during the writing process, such as *planning* (e.g., Flower & Hayes, 1981a; Jones & Tetroe, 1987; Manchón & Roca de Larios, 2007; see also Hayes & Gradwohl Nash, 1996 for an overview of early research), *formulating* (e.g., Roca de Larios, Manchón, & Murphy, 2006; Roca de Larios, Marín, & Murphy, 2001), or *revising* (e.g., Broekkamp & Van den Bergh, 1996; Stevenson, Schoonen, & De Glopper, 2006; Thorson, 2000). However, to provide more specific advice for writing instruction more detailed research is required on the orchestration of specific cognitive activities and how their occurrence is related to text quality (Breetvelt, et al*.,* 1994, p. 104).

In addition, some studies on process-product relations (i.e., relations between the occurrence of cognitive activities and text quality) focused on the role of task representation during the writing process. The task representation that a writer forms at the start of the writing process changes under the influence of the cognitive activities that are carried out while writing, such as reading the assignment, reading the text produced so far, and generating (Carey & Flower, 1989; Van der Hoeven, 1996a). This means that writers must take into account *"the changing task situation during writing"* (Van den Bergh & Rijlaarsdam, 2001, p. 374), and continuously adapt the way they apply each cognitive activity. For example, if writers reread the assignment while writing and realize that they misunderstood it in some way, then they must set new goals or make a plan to solve this problem. Moreover, adapting to the task at hand in this way, appears to have a positive influence on text quality (Van Weijen, Van den Bergh, Rijlaarsdam, & Sanders, 2005a).

In line with earlier research, we use the term *orchestration of cognitive activities* to describe the temporal distribution of cognitive activities during the writing process (e.g., Alamargot & Chanquoy, 2001; Braaksma, Rijlaarsdam, Van den Bergh, & Van Hout-Wolters, 2004, p. 2; Graham & Harris, 2000; McCutchen, 2000; McCutchen, Covill, Hoyne, & Mildes, 1994). Research has shown that the temporal distribution of each activity (i.e., the moment at which it occurs during the writing process) can explain variations in text quality between writers (Breetvelt, et al*.,* 1994; Rijlaarsdam & Van den Bergh, 1996, 2006; Roca de Larios, et al., 2001;

Van den Bergh & Rijlaarsdam, 1999, 2001). So, the way the different cognitive activities are carried out during the writing process varies between writers. They differ not only in the amount of attention that is paid to each activity, but more importantly, in the moment at which they execute each activity. For instance, Breetvelt, et al. (1994) found that writers who were more likely to read the assignment at the start of the writing process were more likely to produce good quality texts than those who were less likely to focus on reading the assignment at that time. In addition, they found that writers who were more likely to focus on formulating near the end of the writing process were more likely to produce good quality texts than those who did not (see Breetvelt, et al., 1994). Another example is a study by Van den Bergh and Rijlaarsdam (1999) on the relationship between different types of generating and text quality. They showed that this correlation changes under the influence of the activity that generating is preceded by. For example, when generating is directly preceded by reading the assignment, the correlation with text quality is strongest at the start of the writing process and then gradually decreases. But the opposite occurs when generating is preceded by reading the text produced so far: then the correlation with text quality is negative in the beginning but becomes positive near the end of the writing process. So, variations in the orchestration of these cognitive activities seem related to variations in text quality *between* writers (see also Rijlaarsdam & Van den Bergh, 2006). However, whether writers vary the orchestration of cognitive activities between task, and whether such variations are also related to variations in text quality *within* writers is as yet unclear.

Most studies on process-product relations in the past share a single methodological problem: Writers were usually required to write only one text each (or sometimes two, Breetvelt, et al., 1994; Van der Hoeven, 1996a). This is problematic for two reasons. First of all, it is impossible to determine which part of the variation in process-product relations can be attributed to interwriter variation (i.e., between writers) or to intrawriter variation (i.e., variation between tasks), which is also present. These two variation components can only be determined by estimating the intrawriter variation separately, using a multiple-task-per-participant design. Second, single-text-designs are also problematic, because research has shown that text quality varies strongly within writers (Coffman, 1966; Schoonen, 1991; Wesdorp, 1974). To adequately determine writing ability on an individual level, multiple texts by the same author must be assessed by several judges. Estimates of the number of texts that must be examined vary from four (Coffman, 1966) to twenty texts (Van den Bergh, De Glopper, & Schoonen, 1988), depending on the type of task, ranging from general essays to more specific functional texts. Thus results from studies which only examined a single text per writer are hard to interpret, and the generalizability of their results over tasks is seriously limited.

Levy and Ransdell's studies form an exception, in several ways (Levy & Ransdell, 1995, 1996). Their participants were required to write a text each week, over a period of 10 to 12 weeks, which enabled them to study writer's behaviour across tasks. They found that writers' behaviour was so stable across tasks for individual writers, that they could be described as *"writing signatures"* (Levy & Ransdell, 1996, p. 149). But their design and method were rather different from those of other studies mentioned above. Their coding scheme, for example, was far less fine-grained than that of other studies (e.g., Breetvelt, et al., 1994; Van den Bergh & Rijlaarsdam, 1999, 2001; Van der Hoeven, 1996a). It only contained four categories: planning, text generating, revising and reviewing, while the schemes of other studies usually contained eleven or more categories. This means that we cannot determine whether the stability described by Levy and Ransdell (1995, 1996) is due to their chosen research design or whether those other studies would also have found stable writing behaviours if more tasks had been carried out.

Therefore, in this study, we wish to determine whether variation in process-product relations within writers exists (between writing tasks) by asking writers to write multiple texts. By using a design and method of analysis similar to that of earlier studies, we can investigate whether earlier findings based on a single task per participant (e.g., Van den Bergh & Rijlaarsdam, 1999, 2001) are generalizable. We will try to establish whether writers vary the orchestration of the writing process (i.e., the way they employ the different cognitive activities) across tasks or whether this behaviour is rather stable across tasks (see Levy & Ransdell, 1995, 1996). We already know that their orchestration varies *during the writing process* (see Breetvelt, et al., 1994; Van den Bergh & Rijlaarsdam, 2001). But the question is whether it also varies *between* tasks. And if so, what is the effect of such variation on text quality? Does varying the orchestration of cognitive activities between writing tasks result in relatively good quality texts? Or is it a sign of inexperience and does it result in relatively weak texts?

Research questions

This study focuses on the variation in the orchestration of cognitive activities and text quality both between and within writers. The main questions are: *(1) Do writers vary the orchestration of each cognitive activity during the writing process between tasks?* and *(2) Does the relation between the orchestration of activities and text quality vary between tasks?*

In line with earlier research, we assume that the temporal distribution of each activity will change over time during the writing process (see Breetvelt, et al., 1994; Van den Bergh & Rijlaarsdam, 2001). With regard to the first research question, we predict that writers will vary the orchestration of cognitive activities between tasks, although some writers may vary their orchestration more between tasks than others. Regarding the second question, and in line with earlier research, we predict that the relationship between the occurrence of each cognitive activity and text

quality will vary over time. Activities which correlate positively with text quality during one phase of the writing process might not correlate with text quality during other phases. In addition, more specific for the multiple-writing-task design, we predict that the relationship between cognitive activities and text quality will vary between tasks (within writers) as well.

METHOD

Participants

The participants in this study were twenty first-year English students. Their average age was 18 years and the majority was female (85%)[2]. At the time of the study they were all enrolled in the first-year program of the English Department. We selected BA English students for this study because we wished to include a fairly homogenous group of participants and because there was a large population of first-year students to choose from. Participants all received a small financial reward for their efforts.

Procedure

Participants wrote four short argumentative essays each in their L1 (Dutch) under think-aloud conditions. They each wrote their texts during four individual sessions, over a period of roughly twelve weeks. Each writing session was recorded with audio and video equipment. Participants were told that they had 30 minutes to complete each assignment, but they could stop when they thought they were done. As a result the time writers needed to complete their tasks varied enormously, from 10.79 minutes to 37.64 minutes, while the average writing time was 23.46 minutes (sd = 7.23 minutes) per session. However, no effect was found of task order on writing time (F (3, 76) = .81; p = .49). During the first session, each participant was trained to work under think-aloud conditions. They were first given a brief explanation of what thinking aloud entails:

> *We would like you to think aloud while writing your essay. That means that you try to say aloud everything that you think while writing. Thinking aloud may seem a bit strange at first, but once you start you'll notice that you get used to it quite quickly.*

Then they practiced thinking aloud, while solving an algebra puzzle, a small crossword puzzle and while writing a short text (roughly five lines) on one of ten randomly assigned topics (e.g., their favourite food, book or pet, or what they did

[2] The high percentage of female participants in this study is a reflection of the male-female ratio of the student population in the English department.

last weekend etc.). Before starting each subsequent session, they each wrote another short text, similar to the first one but on a different topic, to remind them how to think aloud. The experimenter was present during each session and participants were prompted to continue thinking aloud if they fell silent.

Although there has been much discussion regarding the pro's and con's of using the think-aloud method, research has indicated that the possibly detrimental influence of verbalizing on the cognitive activities that take place is not as large as is often assumed (see for example, Ericsson, 1998; Janssen, Van Waes, & Van den Bergh, 1996; Manchón, Murphy, & Roca de Larios, 2005; Roca de Larios, et al., 2006). In addition, concurrent verbalizations appear to provide the most direct indication of a writer's thought processes and also seem *"more valid than other forms of verbal accounts because of the absence of a time interval between the occurrence of a thought and its expression"* (Roca de Larios, et al., 2006, p. 104; see also Ericsson, 1998; Hayes & Gradwohl Nash, 1996, p. 45). Therefore, it still seems the best way, at present, to observe the occurrence of cognitive activities during the writing process.

Assignments

Each participant wrote four essays, on four of eight possible topics[3] included in the study. These eight assignments were identical except for their topic; all other task features were kept constant. The basic assignment was to write a short essay for a fictitious essay contest in the local University Newspaper. It consisted of information on the task, such as the goal, topic, and target audience of the assignment and six short sources on the topic, at least two of which had to be included in the text. Participants were asked to give their opinion and argue for or against the topic of the assignment. The topics were assigned to the participants based on their participant number. Participants 1 through 19 (odd numbers) were assigned the first four topics: *Education in English*, *Surveillance Cameras*, *Having Children,* and *Mobile Phone Use*, while Participants 2 through 20 (even numbers), were assigned the other four topics: *Compulsory Organ Donation, Downloading Music*, *Life as a Student,* and *Soft Drug Legislation* (see Appendix A). The tasks were subsequently executed by each participant in random order.

Data

The audio and video recordings of each writing session were transcribed and turned into think-aloud protocols. The protocols were subsequently segmented and coded to distinguish between eleven different cognitive activities (see Table 1) and three extra categories. The extra categories were: interactions between the participant and the researcher, physical actions, such as scratching one's nose, taking a sip of water or turning the page, and actions concerning the control of the

[3] There were eight topics because students also wrote four texts in their L2 (English).

computer, such as mouse movements and the use of the enter, tab and arrow keys. The procedure for segmenting and coding the protocols was based on Hayes and Flower (1980), and was also used in previous research (e.g., Breetvelt, et al., 1994; Couzijn, Van den Bergh, & Rijlaarsdam 2002; Van den Bergh, 1999; Van den Bergh & Rijlaarsdam, 1999; Van der Hoeven, 1996a).

Table 1: Overview of the eleven main categories included in the coding scheme

Cognitive activity	*Description and protocol example*
(Re)-Reading the assignment	Reading the assignment and/or the sources
Planning: Self-instructions	Instructions the participant gives himself regarding the next step in the writing process (e.g., *"I need to decide how to start"*)
Planning: Goal setting	Paraphrasing task demands in the assignment or setting new goals (e.g., *"This text is going to be about..."*)
Planning: Structuring	Selecting and ordering ideas or making a rough outline (e.g., *"First the pro's, then the con's and then my own opinion!"*)
Generating ideas	Paraphrasing ideas from the assignment or generating new ones: (e.g., *"Us humans we're very curious and we like to gossip!"*)
Metacomments	Reflections on the writing process as a whole or comments on the assignment and sources (e.g., *"I had the perfect sentence in my head"*)
Pausing	Silence or interjections (e.g., *um, ehhh*)
Formulating	Text production or typing the actual words on the computer.
Rereading own text	Rereading short or longer segments of the text produced so far
Evaluating own text	Evaluating the text produced so far at word, sentence or text level (e.g., *"No I'm not gonna say this"*)
Revising own text	Revising the text produced so far at word, sentence or text level (e.g., erases 'them' and types 'people')

The protocols were segmented by a single researcher so that each protocol segment contained a single occurrence of a cognitive activity. In other words, every time a writer switched from one activity to the next (e.g., from Planning to Formulating) this new activity was placed in a new segment (see Breuker, Elshout, Van Someren, & Wielinga, 1986). Subsequently, the protocol segments were coded by six different coders. The interrater reliability between coders proved satisfactory (Cohen's kappa = .84). Due to the rather fine-grained segmentation applied in this study, the

average length of protocol segments, over writers and tasks, was 5.4 words (*sd* = .98). This varied between protocols from 3.75 words to 8.70 words.

For the purpose of this study, the analysis of the writing process will focus on four cognitive activities: *Reading the assignment, Planning, Generating ideas*, and *Formulating*. Reading the assignment is defined as any occurrence of reading the assignment and the sources which occurred out loud. Reading was included in the analysis because it is usually the first activity writers engage in after they are provided with an assignment by their teacher or a researcher. Furthermore, research has shown that writers' reading behaviour varies when they are required to incorporate sources in their texts (Kennedy, 1985). Planning was included in the analysis because it is one of the most influential activities in the writing process (see for example Breetvelt, et al*.,* 1994; Flower & Hayes, 1980a, 1981a; Levy & Ransdell, 1995; Van den Bergh & Rijlaarsdam, 2001). We included three types of planning in this category: self-instructions, goal setting, and structuring, which are all activities that guide or direct the writing process in some way (see also Hayes & Gradwohl Nash, 1996, p. 33). Generating ideas is classified as a separate activity, because it is most directly related to the content of the text and does not influence the other activities to the same extent as Planning does. It was included in the analysis because it is a central part of the writing process: *"If one has no ideas, or has no access to ideas, one has nothing to write about"* (Van den Bergh & Rijlaarsdam, 1999, p. 99). Finally, Formulating is defined as text production, which includes the construction of words and sentences as well as the act of typing on the computer. It was selected because it forms the core of the writing process (see for example McCutchen, et al., 1994, p. 256) and also because it has received relatively little attention in previous research (see Roca de Larios, et al*.,* 2001). Revision is not included in the analysis, because the majority of revisions in this study consisted of minor editing actions at word-level, which are less interesting or informative than other cognitive activities.

The focus of this study is on the distribution of cognitive activities over the writing process instead of on each activity's frequency of occurrence. However, the average frequency of occurrence can provide a general indication of differences between writers and tasks. Therefore, the average frequencies of occurrence for all four activities, averaged over writers and tasks (N = 80), and the average number of segments in the writing process as a whole (i.e., in each think-aloud protocol) are presented in Table 2. In general, Formulating occurs far more frequently (in 17.2% of segments) than Generating (5.3%), Reading the assignment (3.7%), and Planning (3.1%). The frequency of occurrence of each activity clearly varies between writers and tasks; the variation is particularly large for Generating and Formulating (see range and standard deviations in Table 2). In addition, note that Self-instructions dominate the Planning category (81%), while Structuring (15%) and Goal setting (4%) occur far less often.

Table 2: Average frequencies of occurrence and range for all four cognitive activities over all texts and writers (N = 80)

Cognitive activity	*Average number of segments (sd)*	*Range (number of segments)*	
		Min.	*Max.*
Reading the assignment	26.46 (11.28)	10	77
Planning: Total	22.50 (16.92)	0	98
Planning: Self-instructions	18.21 (12.74)	0	62
Planning: Goal setting	0.88 (1.93)	0	13
Planning: Structuring	3.41 (5.36)	0	28
Generating ideas	38.05 (36.53)	0	187
Formulating	123.9 (44.52)	45	234
Writing process as a whole	721.53 (290.41)	222	1529

Text quality

Each participant wrote four texts (*N* texts = 80). Texts ranged in length from 137 to 475 words, with an average length of 292 words (*sd* = 79 words). The order in which the texts were produced did not have a significant effect on text length (F (3, 76) = .03; p = .99). The quality of the 80 texts was assessed twice: analytically and holistically, by the same jury of five judges. The analytical rating was performed first, using an analytical rating form, with four main categories: *Structure, Content, Argumentation,* and *Conclusion* (see Appendix B). The rating form was initially based on a form used in earlier research (see Breetvelt, et al*.,* 1994) and was adapted to increase the focus on structure and argumentation (see Sanders & Schilperoord, 2006). We expected structure and argumentation to be text features which were likely to be stable between texts by the same writer.

The holistic rating was performed roughly two weeks later. Judges were provided with a benchmark essay on each topic worth the 'average' score (100 points). They had to decide how much better or worse each text was than the average text. To help them with their rating, they were also provided with an example of how essays should generally be structured (see Appendix C).

The within jury reliability proved satisfactory for both methods (analytical: α = .88, holistic: α = .82). The correlation between the two methods proved quite high (r = .74). Corrected for attenuation the correlation between the two scores was .87. Therefore the scores were combined into a single general score for text quality on an arbitrary standardized scale. The average scores per topic and the average score for all tasks are shown in Table 3. They are presented as z-scores, to facilitate comparison between tasks. Although the average scores are not the same for each topic, there was no evidence of order effects on text quality scores (F (3,

76) = .40; p = .75). To provide an indication of the type of essays produced, an example of an average and a good text by two different writers are presented in Appendix D.

Table 3: Average text quality and range per topic in z-scores (N = 80)

Task	*Topic*	*N*	*Mean (sd)*	*Min.*	*Max.*
A	*Education in English: good idea or bad suggestion?*	10	-0.16 (0.87)	-2.05	1.12
B	*Having children, yes or no?*	10	0.12 (1.27)	-1.76	1.51
C	*Everybody should automatically be a donor, unless they explicitly register against it: good or bad idea?*	10	0.06 (0.95)	-1.10	1.79
D	*Life as a student: A hard life full of stress? Or isn't it that bad?*	10	-0.25 (1.19)	-1.89	1.43
E	*Do surveillance cameras in inner city areas increase public security?*	10	0.14 (1.09)	-1.96	1.69
F	*Mobile phones: irritating or essential?*	10	0.30 (1.02)	-1.59	1.28
G	*Downloading music for free: criminal or should be possible?*	10	-0.17 (0.77)	-1.63	0.86
H	*Soft drugs: should they be legalised or prohibited completely?*	10	-0.03 (1.02)	-1.64	1.65
Total		80	0.00 (1)	-2.05	1.79

Method of analysis

This study was set up to investigate whether writers vary the orchestration of cognitive activities between tasks, and if so, whether variation between tasks is related to text quality. The data are hierarchical in nature, as the observations of activities (protocol segments) are nested within tasks, which are themselves nested within participants. Therefore a multilevel analysis was performed on the data. This type of analysis enables us to model the occurrence of a cognitive activity as a function of time during the writing process. For such, in essence longitudinal models, several types of models have been suggested (see Breetvelt, et al., 1994). We prefer polynomial models, that is, the occurrence of a given activity is modelled as a function of the moment at which it occurs (Van den Bergh & Rijlaarsdam, 1996). Such polynomials (i.e., powers of the moment) are extremely flexible and

can take almost any shape depending on the number of powers included in the model and their numerical values. If each segment Y is coded according to whether a given activity does ($Y_{ijk} = 1$) or does not occur ($Y_{ijk} = 0$) at moment i ($i = 1, 2, ... i_{jk}$) in task j ($j = 1, 2, 3, 4$) of writer k ($k = 1, 2, ... 20$), we can write the model as:

$$Logit(Y_{ijk}) = \beta_{0\,jk} * M^{0}{}_{ijk} + \beta_{1\,jk} * M^{1}{}_{ijk} + \beta_{2\,jk} * M^{2}{}_{ijk} + ... \qquad (1)$$

We can include as many powers of the moment ($M^0, M^1, M^2, \ldots$) in the polynomial as needed to accurately describe the observed data. A general rule of thumb is that higher order coefficients are not included in the model if lower order coefficients lack significance. The regression weights in the model ($\beta_0, \beta_1, \ldots$) are indexed with j and k, showing that they can vary between tasks (j) as well as between participants (k). We can write the regression weights as:

$$\begin{aligned} &\beta_{0\,jk} = \beta_{000} + u_{0\,jk} + v_{00k} \\ &\beta_{1\,jk} = \beta_{100} + u_{1\,jk} + v_{10k} \\ &.... \end{aligned} \qquad (2)$$

That is, for the mean regression weights ($\beta_{000}, \beta_{100}, \ldots$) the variance between tasks ($S^2 u_{0jk}, S^2 u_{1jk}, \ldots$) and a variance component between individuals ($S^2 v_{00k}, S^2 v_{10k}, \ldots$) are estimated. The variance between tasks indicates the degree to which the occurrence of an activity varies between tasks. If this coefficient is low or non-significant, there are no substantial differences between tasks. The second type of coefficient indicates the differences between individuals. The residuals for the individual writers ($v_{00k}, v_{10k}, \ldots$) represent deviations for the individual writer from the overall mean. The task and writer specific residuals ($u_{0jk}, u_{1jk}, \ldots$) represent the deviations for task j from the mean of writer k. Inserting the second equation in the first gives the model to be estimated:

$$\begin{aligned} &Logit(Y_{ijk}) = M^{0}{}_{ijk}(\beta_{000} + u_{0\,jk} + v_{00k}) + \\ &M^{1}{}_{ijk}(\beta_{100} + u_{1\,jk} + v_{10k}) + \\ &M^{2}{}_{ijk}(\beta_{200} + u_{2\,jk} + v_{20k}) + \\ &.... \end{aligned} \qquad (3)$$

The fixed part of the model ($\beta_0, \beta_1, \ldots$) represents the average change in occurrence of each activity during task execution. The random part shows the differences due to individuals and due to task. Equation 3 shows that the variance between tasks as well as the variance between individuals (the random part of the model) depends

on the moment. That is, the variance between tasks at moment M, can be approximated as:

$$S^2_{task} = S^2_{u0\,jk} + 2 * S_{u0\,jk,u1\,jk} * M^1_{ijk} + S^2_{u1\,jk} * M^2_{ijk} + \ldots \quad (4)$$

The same type of formula can be used for the variance between individuals, except that the residuals denote the deviations for individual *j* (v_{00k}, v_{10k}, ...). Both variance estimates together determine the intraclass correlation, which is defined as the between-writer variation divided by the total variation (Snijders & Bosker, 1999). This coefficient shows the proportion of variance related to differences between individuals. Hence, this coefficient can also be used to show the influence of tasks on the writing process. Please note that the variance (at each level) is a continuous function of the moment (M) during the writing process. Hence, the heteroscedasticity of the variance depends on the significance of the individual parameters. This can be tested by using a t-value (i.e., [estimate/s.e.] > 1.965).

The relation between each activity and text quality is estimated in a second model. In essence, this comes down to estimating the correlation between the residual scores for task *j* (u_{0jk}, u_{1jk}, …) as well as the correlation between the residuals for individuals (v_{00k}, v_{10k}, …) and text quality. Please note that these correlations depend on the value of M_{ijk} and its subsequent powers (see Van den Bergh & Rijlaarsdam, 1996). In this analysis, the segment number was used as an indication of the moment (Van den Bergh, Rijlaarsdam, Janssen, Braaksma, Van Weijen, & Tillema, in press). This is possible, because there was a very high correlation between the number of protocol segments and total writing time (r = .89). In other words, the longer writers took to complete their writing tasks, the more segments their protocols contained (see also Braaksma, et al., 2004).

RESULTS

This study was set up to answer two main questions: (1) *Do writers vary the orchestration of each cognitive activity during the writing process between tasks?* and (2) *Does the relation between the orchestration of activities and text quality vary between tasks?* To answer the first question, we must first determine the average temporal distribution of each cognitive activity, over tasks and writers, and then the interwriter variation (i.e., how each writer's average behaviour deviates from this average). Subsequently, we must calculate the intrawriter variation, which indicates how writers vary the orchestration of activities between tasks, compared to their own average. To answer the second question, we will examine the relationship between the inter- and intrawriter variation of cognitive activities and text quality, by calculating the correlation between the occurrence of each activity and text quality. Then we can

determine whether the relation between the orchestration of each activity and text quality varies between tasks.

Parameter estimates of the model are hard to interpret, as they are presented in logits[4] (see Appendix E), a nonlinear transformation of proportions (see Equations 1-3). Therefore we will discuss the results converted to proportions that indicate the probability that an activity will occur at a specific moment during the writing process. These results, as presented in Figures 1 through 4, summarize the 57728 occurrences of cognitive activities over 4 tasks for each of the 20 participants (see Table 2). The specific model used to estimate the orchestration of an activity during the writing process differs for each activity, in the number of parameters included (i.e., the number of powers of the moment, see Appendix E). On average, the model fits the data quite well. In general, the proportion of activities classified correctly exceeded 92% (Reading: 94.8%, Planning: 96.8%, Generating: 92.9%, Formulating: 94.1%).

Please note that in some figures the curves are not all of equal length, due to the fact that participants differed in the time they needed to complete each writing task, and thus the number of segments in each think-aloud protocol varied accordingly. In addition, as the segment number (x-axis) increases, the number of observations decreases, because only a few writers wrote for a very long time, and thus had a large number of segments (above 1000). Any results from segment 1000 onwards are therefore based on the writing processes of only a few participants and must be interpreted with caution.

Average temporal distributions and interwriter variation

Figure 1 shows the average curves for the probability of occurrence of the four activities, averaged over writers and tasks[5]. In general, participants had a higher probability of occurrence for Formulating than for the other three activities. In other words, they are more likely, in general, to spend time Formulating than Reading, Planning, or Generating. The exceptions are the start and end of the writing process, when writers are most likely to focus on Reading the assignment. In addition, the probability of occurrence for all four activities varies over time. Reading is most likely to occur at the start and end of the writing process. Planning and Generating are most likely to occur in the beginning, whereas Formulating is most likely to occur in the middle of the writing process. Furthermore, the probability of occurrence for Generating and Planning are similar and very small, although Generating is slightly more likely to occur overall than Planning.

The interwriter variation for each activity, generalized over tasks, are shown in Figure 2. The four panels in the figure each present the average curves

---

[4] Please note: logit (P) = ln ($e^p/(1+ e^p)$).

[5] Please note that the scales on the Y-axis of the two panels are different, due to the large differences in the probability of occurrence between the two pairs of activities.

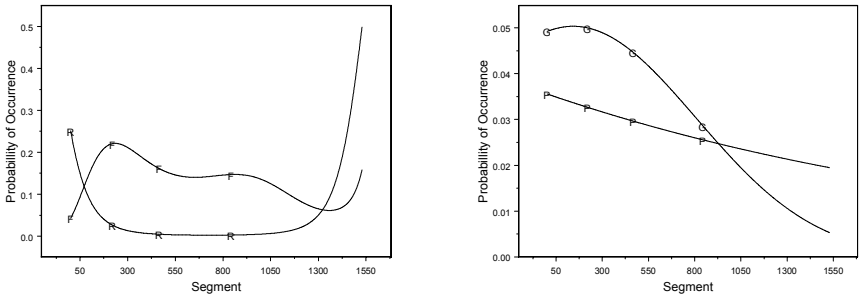

Figure 1: The mean change in the probability of occurrence of Reading the assignment (R) and Formulating (F) (left) and Planning (P) and Generating (G) (right).

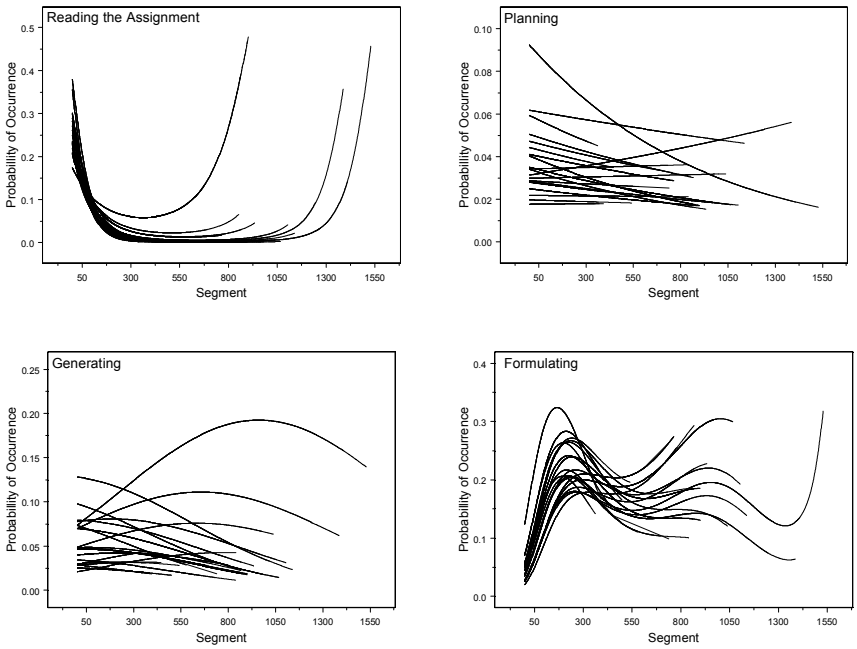

Figure 2: The average probability of occurrence for each cognitive activity per writer (n = 20).

per writer (over four tasks) for a specific cognitive activity. Thus each figure contains 20 curves, one for each individual writer. The results presented in Figure 2 show two things. First of all, they indicate the extent to which the orchestration of each activity varies on average between writers. If differences between writers are small, then the 20 curves will be practically identical, but if they are large, then the curves will fall further apart. Second, the results in Figure 2 indicate to what extent the orchestration of each activity varies on average over time per writer. If there is little or no variation then the curves will be almost horizontal, whereas if there is a lot of variation over time, they will vary in shape.

Overall, results indicate that each activity varies both between writers and over time (during the writing process), although results are more apparent for some activities than for others. Reading the assignment, for example, is most likely to occur at the start of the writing process and in addition, the between writer variation is rather large at that point (range: .17 to .38). Subsequently, the differences between writers decrease and it becomes almost impossible to distinguish between them (minus a few exceptions). In general Planning has a low probability of occurrence and the interwriter variation appears somewhat small (range: .01 to .09). On the whole, participants are most likely to plan in the beginning, but some do the reverse and are more likely to plan as the writing process unfolds. For Generating, the variance between writers appears small initially but then increases over time and becomes largest in the middle of the writing process (range: .03 to .18). For most participants, the probability of occurrence is highest in the beginning and decreases as the writing process continues, although some appear to do the reverse. Formulating is most likely to occur on average after the start of the writing process. The interwriter variation appears small initially, but the differences increase as the writing process unfolds (see Figure 2). In addition, writers' Formulating behaviour clearly varies over the writing process. Writers' Formulating curves appear to differ: in time spent on task (curve length), in their probability of occurrence (curve height), and in the flow of their writing processes (one high peak or several smaller ones).

Intrawriter variation: differences in process execution between tasks

The next step is to examine the variation in the occurrence of each activity between tasks within writers (see Appendix E for estimates). Therefore we calculated the average probability of occurrence for each cognitive activity per writer as well as the 80% confidence intervals. These intervals indicate how much a writer's behaviour for a specific activity varies between tasks. In other words, there is an 80% chance that, if presented with a similar type of task, a writer's behaviour will fall within these outer limits. A considerable distance between the 80% confidence intervals and the mean indicates that the intrawriter variation is large: The writer's behaviour varies strongly over tasks. If the distance between the confidence

intervals and the mean is small, the writer's behaviour is rather stable over tasks. To illustrate the intrawriter variation we have chosen to provide a figure representing the average variation (over all writers) and a figure representing the intrawriter variation for a specific writer for each cognitive activity. In each case we chose a writer with relatively high intrawriter variation for that activity (see Figure 3).

On average, the probability of occurrence for Reading varies within writers and, more importantly, also varies during task execution (see Figure 3, top row, left). In general, the differences between tasks within writers are large, especially in the beginning and near the end of the writing process, when the differences within writers (between tasks) are much larger than the differences between writers (compare Figures 2 and 3).

To further illustrate the intrawriter variation for Reading, we included the results for a specific writer, Writer 17 (see Figure 3, top row, right). This writer's Reading behaviour clearly varies over tasks at the start of the writing process, but subsequently the variation decreases rapidly, most notably between Task E and the other three tasks. During Task E the probability for Reading was rather high throughout the writing process and increased towards the end, while during the other three tasks this writer only focused on Reading the assignment at the very start of the writing process. In other words, the intrawriter variation for this writer is strongly influenced by her behaviour on Task E.

Although Planning usually has a very low probability of occurrence (see also Figures 1 and 2), its probability of occurrence clearly varies over time and between tasks. The intrawriter variation is largest, on average, in the beginning and near the end of the writing process (see Figure 3, second row, left). Please note that the increase in variation at the start and near the end of the writing process is not necessarily due to the same task. The upper and lower confidence intervals indicate the interval within which the variation between tasks falls, but we cannot accurately predict how each activity is actually carried out during a specific task. It is possible that for one task the probability of occurrence for Planning decreases over the writing process (from high at the start to low near the end of the writing process), while for another task it decreases after the start of the writing process but subsequently increases again after the middle of the writing process.

The results for Planning are illustrated by the intrawriter variation for Writer 8 (see Figure 3, second row, right). For this writer, the variation between tasks is largest at the start of the writing process, after which the differences between tasks gradually decrease. The variation is rather small between Tasks C and H and largest between Tasks D and G. The probability of occurrence for Planning appears to decrease as the writing process unfolds for Tasks C, D, and H, but increase slightly over time during Task G. So this writer's Planning behaviour appears to vary somewhat under influence of the task at hand. The increase in

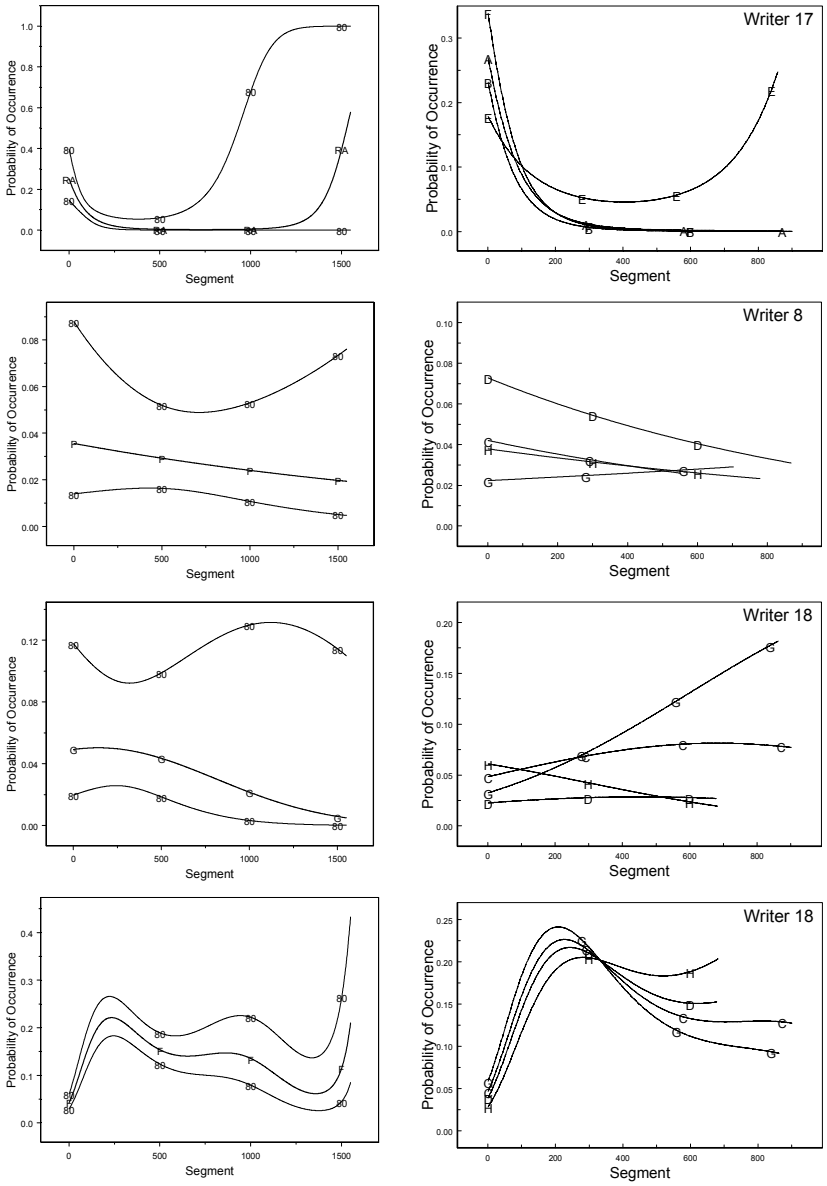

Figure 3: The intrawriter variation for each cognitive activity: Average probability of occurrence and 80% confidence intervals (80) (left) and intrawriter variation for one specific writer (4 tasks) (right) (from top to bottom: Reading the assignment (RA), Planning (P), Generating (G), and Formulating (F)).

Planning near the end of the writing process could, for example, be related to problems the writer encountered while writing. When problems occur, writers use self-instructions (a sub-process of Planning in this study) to keep themselves on track and to solve their problems.

Generating is somewhat more likely to occur, on the whole, than Planning. Although it is not a very frequent process, the intrawriter variation is rather large, in general, throughout the writing process (see Figure 3, third row, left). It is particularly large at the start and end of the writing process, although the variation at the end is not very reliable due to the decrease in the sample over time, as described earlier. This is clearly illustrated by the intrawriter variation for Writer 18 (see Figure 3, third row, right). This writer has some variation between tasks at the start of the writing process, but the variation becomes extremely large as the writing process unfolds. This is mainly due to Task G, during which the probability of occurrence for Generating increases enormously over time. So, near the end of the writing process this writer's Generating behaviour varies strongly under the influence of the task at hand; in two tasks (D and H) she is unlikely to generate new information near the end, while she is much more likely to do so in her other tasks (C and G).

The variation between tasks for Formulating is relatively small initially, but increases during the middle of the writing process (see Figure 3, bottom row, left). Writers are likely to distribute their Formulating activities in roughly the same way across multiple tasks. This is confirmed by the intrawriter variation for Writer 18 (see Figure 3, bottom row, right). The variation between tasks is rather small for this writer in the beginning, after which the variation between tasks increases over time. So, writers' Formulating behaviour appears relatively stable over tasks. Overall, there is variation between tasks for all four cognitive activities, although the size of the variation is different for each activity and also varies during the writing process.

Proportion of task-related variation

To determine the extent to which the overall variation is due to tasks, we calculated the intraclass correlations for each cognitive activity, which indicates the proportion of the total variation due to individuals. Subtracting the intraclass correlation from 1 gives the proportion of the variation due to tasks. So, if the values in Table 4 are high, then the variation in the orchestration of each cognitive activity is mainly due to between-writer variation. But if the values are low, then there is mainly task-related variation.

Table 4 shows that in general, the proportion of variation due to individuals is smaller than the proportion of variation due to task, which indicates that most of the variation is due to task effects. This holds for example for Reading the assignment and Planning, although it fluctuates somewhat over time. For

Generating, the between task variation is largest near the end of the writing process, while for Formulating it is largest in the middle of the writing process, after which it decreases rapidly near the end of the writing process.

Table 4: The intraclass correlations indicating the proportion of variation between writers

	Proportion of variation between writers		
Cognitive activity	*Start*	*Middle*	*End*
Reading the assignment	.23	.43	.44
Planning	.40	.48	.40
Generating	.50	.58	.40
Formulating	.40	.23	.80

By and large, it seems that the proportion of variation between individuals depends on the cognitive activity being carried out as well as on the moment during the writing process. The same applies for the proportion of variation due to task. The between task variation can be very large, depending both on the activity taking place as well as on the moment during the writing process. This is the case, for example, for Reading the assignment, for which the between task variation is very large near the end of the writing process (.56), although this is probably due to some extent to the fact that the number of writers in the sample decreases over time. This means that the occurrence of Reading depends greatly on the task being carried out, not on the individual. Thus differences in patterns of orchestration for each cognitive activity can be attributed more to tasks than to individual writers. Furthermore, as the differences between tasks are much larger, in general, than the differences between individuals, generalisations over tasks, based on a few tasks per writer becomes almost impossible.

Variation in the relation between cognitive activities and text quality

To answer research question two, we needed to determine whether the relation between the orchestration of cognitive activities and text quality varies between tasks. However, the relation between each activity and text quality can also vary during the writing process. Therefore, we first estimated the average correlation (over writers and tasks) between the temporal distribution of each activity and text quality (see solid lines in Figure 4). This tells us how the correlation between each activity and text quality varies on average over time. Second, to determine to what extent this correlation varies from task to task, we also calculated the 80%

confidence intervals due to task differences. These confidence intervals indicate the range within which other tasks are likely to fall. Thus if the lines fall far apart, the correlation is likely to vary greatly between tasks, whereas if they fall close together, then the correlation is likely to be relatively stable between tasks. We will first discuss to what extent the relation between each cognitive activity and text quality varies over time, and then discuss how this correlation varies between tasks for each activity.

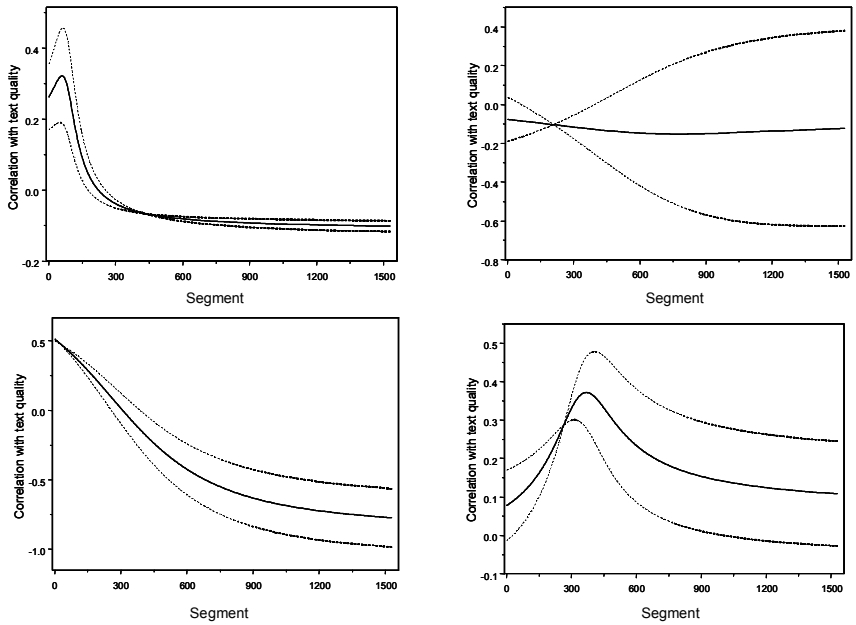

Figure 4: The correlation between Reading (top left), Planning (top right), Generating (bottom left), and Formulating (bottom right) and text quality: average (solid line) and 80% confidence intervals due to task (dotted lines).

The correlation between Reading the assignment and text quality varies during the writing process from .33 to -.10 (solid line, see Figure 4). The correlation is highest in the early stages of the writing process (± .27 to .33), although it is not extremely strong. This suggests that students who show a high probability of occurrence for Reading the assignment in the beginning, on average, are likely to have written better texts than those whose probability was lower. The correlation then drops rather quickly to about -.10, which indicates that students who were likely to read or reread the assignment regularly at later moments in the writing process wrote poorer texts than students who were less likely to do so (for example, the three writers with the very high curves near the end of the writing process in Figure 2).

The correlation between Planning and text quality (solid line) is weak throughout the writing process (if significant at all). It seems best, in general, to plan very early on, after which the correlation with text quality becomes negative. So, participants who have a high probability of occurrence for Planning at the start of the writing process, on average, are likely to have written good quality texts. This holds, for example, for the writers with the highest probability of occurrence at the start in Figure 2. However, some writers do the reverse and have a relatively high probability of occurrence for Planning near the end of the writing process. Those writers are likely to have written poorer quality texts, such as the writer whose curve continually increases over time in Figure 2.

The correlation between Generating and text quality is strong in the beginning (0.5), but subsequently decreases rapidly as the writing process unfolds (solid line) and then becomes negative (see Figure 4). This indicates that writers who have a high probability of occurrence for Generating at the start of the writing process, on average, are likely to have written better texts than those whose probability was lower. This holds for example for those writers with the highest curves for Generating at the start of the writing process in Figure 2. Furthermore, it also indicates that writers who generate many ideas after the start of the writing process are likely to have written texts of poorer quality than writers who did not do so, such as the writer with the highest curve at the end of the writing process in Figure 2.

The correlation between Formulating and text quality increases rather rapidly after the beginning, and then decreases as the writing process progresses (solid line, see Figure 4). This indicates that writers who were likely to concentrate, on average, on Formulating after the start of the writing process wrote better texts than students who were less likely to do so. For instance, the writers with the highest average curves after the start are probably good writers, while the writer with the curve that rises sharply upwards near the end of the writing process in Figure 2 is probably a weak writer.

Overall, the results for these four cognitive activities all confirm that there appears to be a substantial relation between the occurrence of each activity and text quality. However, for each activity this relation varies over time, and can be strong or weak depending on the moment during the writing process.

Variation in process-product relations between tasks

To determine to what extent the relations between the occurrence of cognitive activities and text quality vary between tasks, we must examine the 80% confidence intervals (the dotted lines in Figure 4), as described earlier. Figure 4 shows that the correlation between Reading the assignment and text quality varies initially from .18 to .45 due to tasks. As the writing process continues the confidence intervals clearly become smaller, indicated by the fact that the dotted lines fall rather close to the

average correlation curve (see Figure 4, top left panel). Hence, for Reading, task effects are small, and variations in Reading the assignment due to task have little influence on the relation with text quality during later phases. Furthermore, the estimates (see Appendix E) indicate that there appears to be a positive additive effect[6] at task level. So, writers who are more likely to read the assignment in the beginning than they usually do, are likely to write better texts (see for example Figure 3, Writer 17, Task F). And indeed, Writer 17's text written during Task F was rated as her best text. Near the end of the writing process, the reverse applies; if writers are more likely to read the assignment (above their own average) at the end of the writing process, the resulting texts are likely to be poorer in quality than the other texts they produce. For example, Task E was considered to be Writer 17's poorest quality text by far (see Figure 3, Writer 17, Task E).

For Planning, the correlation between process and text quality depends largely on the task at hand (see Figure 4), as the confidence intervals fall very far from the average curve. The covariance coefficients at task level (see Appendix E) indicate that the correlation between Planning and text quality is only positive in the beginning of the writing process. If writers are likely to plan during later phases of the writing process, the influence of Planning on text quality becomes negative. For example, a writer who is likely to plan relatively often in the beginning and whose Planning behaviour has a steep decrease over time is likely to have written a relatively good text. But if that same writer carries out another task (under similar circumstances) and only has a high probability of planning during a later phase of the writing process, the text produced will be poorer in quality. If we compare this to the results in Figure 3, we see that Writer 8 has one task in which she is clearly more likely to plan at the start of the writing process (Task D) than during her other tasks. Although her Planning behaviour does not decrease very steeply over time, it seems likely that she wrote her best text during Task D (see Figure 3). However, she actually wrote her best text during Task C, while she produced texts of far poorer quality during the other three tasks. Task D resulted in a poor quality text, even though she was most likely to have planned during that task. This could, for example, be due to the effect of other cognitive activities on text quality. In conclusion, Planning behaviour is very task dependent. It is strongly influenced by the interaction between writer characteristics and task characteristics, such as topic knowledge.

Initially, the correlation between Generating and text quality hardly varies between tasks. However, this variation increases rapidly over time (see the confidence intervals in Figure 4). At the end of the writing process the range is

[6] The actual coefficients presented in the appendix are negative. However we centred the segment variable on the mean (segment 432) and divided by 100 (see Appendix E). This means that the sign of the coefficient has to be changed before and after segment 432. So they become positive when they are related to segments in the beginning of the writing process.

rather large (-0.5 to -1.0), which means that Generating only correlates positively with text quality in the beginning of the writing process. Therefore, writers who are likely to generate (relatively) many ideas towards the end of the writing process (above their own average) are likely to produce poorer texts. This means that Writer 18 (Figure 3, third row, right) is likely to have written a better text during Task H, when the probability of occurrence for Generating gradually declines, than during Task G, when it increases strongly over time. Task H did indeed result in the highest quality text, while the other three texts were much poorer in quality. The text produced during Task D was poorest in quality, which could be due to the fact that Generating was hardly likely to occur at all during that task.

Finally, Figure 4 shows that the variation due to task is relatively large for Formulating. For example, for one task the correlation between Formulating and text quality is positive in the beginning whereas for others it is slightly negative. In general, Formulating is positively related to text quality after the start of the writing process. However during later phases differences between tasks become a very important factor, which is indicated by the fact that the confidence intervals (dotted lines) fall very far apart (see Figure 4, bottom right panel). Towards the end of the writing process for instance, the correlation varies from -.03 to 0.28. So if writers are more likely to formulate after the start of the writing process (above their own average) for one task compared to their other tasks, then that is likely to result in a higher quality text. Indeed, Writer 18 (Figure 3, bottom row, right) wrote her best text during Task H, during which the probability of occurrence for Formulating was higher after the beginning than for the other three tasks.

To sum up, the relation between the orchestration of cognitive activities and text quality clearly varies between tasks. Whether this between-task variation is advantageous however (i.e., results in higher quality texts), depends on the moment during the writing process and on the cognitive activity being carried out. For Reading the assignment and Planning, the results appear rather similar. Variation in Reading behaviour between tasks is only beneficial in the initial phase of the writing process. In later phases it no longer matters, as the correlation with text quality becomes roughly zero and the variation between tasks becomes rather small (see Figure 4). The same holds for Planning; variation in Planning behaviour is only desirable in the beginning of the writing process, after which it becomes negative. In addition, it is unclear why the correlation between the occurrence of planning and text quality varies so strongly between tasks from the middle of the writing process onwards (see Figure 4).

The picture for Generating is somewhat different, as variation is only beneficial from the middle of the writing process onwards. This is due to the fact that the variation due to tasks increases after the start of the writing process, while the correlation with text quality decreases (see Figure 4). This indicates that writers,

who are likely to generate more ideas (above their own average) in later phases of the writing process, are likely to produce poorer quality texts.

Finally, for Formulating, variation appears to contribute to text quality in the beginning and the end of the writing process, but not in the middle. Figure 4 shows that the variation between tasks decreases after the start of the writing process, disappears briefly in the middle of the writing process and subsequently increases again near the end of the writing process. In other words, writers who are likely to formulate relatively frequently in the beginning (above their own average) probably write better texts, while writers who are likely to formulate relatively frequently towards the end of the writing process (above their own average) probably write poorer texts.

CONCLUSION AND DISCUSSION

Variation in the orchestration of cognitive activities between tasks

First of all, we have demonstrated that the average probability of occurrence of a given cognitive activity varies over time. The average probability of occurrence is different for all four activities; for Reading the assignment it was highest in the beginning and the end of the writing process, for Planning and Generating it was highest at the start, whilst for Formulating it was highest in the middle of the writing process. These results correspond to those of earlier research (Breetvelt, et al., 1994; Van den Bergh & Rijlaarsdam, 2001). Second, we have shown that the orchestration of each activity varies between writers. This interwriter variation depends both on the activity being carried out and on the moment in the writing process. It can be very large for one activity at moment A, but very small at moment B. For Reading for example, the differences between writers are large in the beginning and the end of the writing process. For Planning and Generating they are moderate throughout the writing process, while for Formulating, the differences are largest in the middle of the writing process. Due to the large differences between tasks, generalisations based on one or two tasks per writer hardly seem warranted.

Third, participants vary the way they orchestrate activities due to the task at hand. Again, this holds for all four activities. Furthermore, differences between tasks are unequally distributed over the writing process. This means that the extent to which writers vary their behaviour between tasks fluctuates over the writing process. For instance for Reading and Planning, task effects are most apparent at the start and end of the writing process, whereas for Generating and Formulating the task effects are largest after the beginning.

In conclusion, there are obvious differences between writers in their orchestration of cognitive activities during the writing process. Each writer,

generalized over tasks, clearly does something different than the other writers. Moreover, individual writers vary the orchestration of activities between tasks. The extent to which their behaviour varies between tasks is different for each cognitive activity and varies from moment to moment during the writing process.

Variation in the orchestration of cognitive activities and text quality

The relationship between the occurrence of each cognitive activity and text quality varies during the writing process. For Reading the correlation is highest in the beginning, but subsequently tails off. It is important to note, however, that although the correlation is .33 at its peak, and thus results in roughly 10% variance explained, this is only at a single moment during the writing process. In actual fact, the correlation is estimated at each separate moment during the writing process, and all these measurements contribute to the percentage of variance explained. Thus the actual percentage variance explained is actually higher than this single correlation of .33 suggests. However, a precise estimate is difficult to give as it fluctuates over time. This holds for all the other activities as well. For Planning, the correlation is fairly stable, but slightly negative, after the start of the writing process. For Generating, the correlation is very strong in the beginning but rapidly inverses as the writing process continues, while for Formulating the correlation with text quality is low in the beginning, rises, and then drops again.

Interestingly, tasks appear to influence the strength of the correlation but not the general pattern, except in the case of Planning. For all tasks the correlation between Reading and text quality is relatively high in the beginning and then levels off. The same conclusion holds for Generating and Formulating, although the differences due to task seem somewhat larger. However, for Planning tasks do appear to influence the pattern of the correlation. For some tasks, Planning is positively correlated with text quality, while for other tasks the correlation is negative.

Breetvelt, et al. (1994) found that the occurrence of self-instructions and goal setting, two of the activities we defined as Planning in this study, correlated positively with text quality near the end of the writing process. At first glance this appears to contradict our results. However, this could be due to the fact that their participants were ninth graders instead of college students. In addition, Breetvelt, et al. did not take into account the results for the two separate tasks in their analysis. We found extremely large intrawriter variation, that is task dependency, in Planning behaviour across four very similar tasks per writer. So it is likely that the variation in Planning behaviour would also increase if the results in Breetvelt, et al.'s study for the two tasks were analyzed separately. Furthermore, the results they found do fall within the upper and lower confidence intervals we found in our study.

Finally, it has been shown that variations in the orchestration of activities, under influence of the task at hand, can be useful, depending on the moment in the

writing process. For Reading and Planning variation is only desirable in the beginning of the writing process. For Generating it is beneficial in the middle of the writing process, while for Formulating variation can be useful in the beginning and the end of the writing process. Therefore, we conclude that variations in the way activities are executed between tasks appear related to variations in text quality, at least for the activities reported in the present study.

Yet, an important question remains: Why do the temporal distributions of these activities vary between tasks? When presented with an assignment, a writer forms a representation of the task at hand. This initial representation varies between tasks under the influence of both task features (e.g., topic, genre, or target audience) and writer characteristics, such as topic or genre knowledge. These factors could make one task easier or more difficult for a writer to carry out than another task. For example, if writers know much about the prescribed topic of the task, then their representations of the task and the goals they set themselves are likely to be different than if they have to write about an unfamiliar topic. Subsequently, topic knowledge is also likely to influence the way the cognitive activities are carried out during the writing process. An unfamiliar topic might call for more re-representations during task execution and thus initiate a different sequence of activities than a more familiar topic.

Further research

Our aims in this study were twofold. First, we wanted to determine whether writers varied the orchestration of different cognitive activities between tasks. Results indicate that the orchestration of each activity does indeed vary within writers, at least for the four activities included in this study. However, whether this is also the case for the other cognitive activities which play a role during the writing process such as revising and evaluating remains to be determined. Nonetheless, we now know that the writing process varies between writers and between tasks (see Table 4) and that the relation between the occurrence of cognitive activities and text quality varies between tasks as well (see Figure 4). Due to this variation in process execution and process-product relations, it seems advisable to collect multiple essays from each writer to measure writers' skill instead of collecting single essays. Furthermore, this seems to explain why earlier studies also advocated the use of multiple tasks to determine writing skill and or text quality (see Coffman, 1966; Schoonen, 1991; Van den Bergh, et al., 1988; Wesdorp, 1974).

However, Levy and Ransdell (1995, 1996) concluded that writers' behaviour was very stable between tasks, while our results indicate that writers' behaviour is far more likely to vary. We suggest that these differences are probably due to the methodological differences between our studies. First, their coding scheme was rather general, as it contained only four main categories, while our scheme contained eleven categories. The fact that they only included four

categories makes the discovery of stable patterns in the transitions between activities far more likely. In addition, they calculated those transitions in a very detailed way. Because text generating was such a high frequent activity in their study, it is likely to have increased the stability of the transitions between the activities. In our study, we used a larger number of coding categories and we did not analyze the transitions between activities in the same way as Levy and Ransdell did. Therefore, it seems unlikely that Formulating, which was also a frequent activity in this study, could have had a similar stabilizing effect on writers' behaviour. Perhaps if we had trained our participants in the same manner and had used only four major coding categories instead of eleven we would also have found more stable behaviour across tasks. Therefore, further research is needed to determine whether the differences between our results and Levy and Ransdell's are indeed due to methodological differences.

Second, we wanted to establish whether the relation between the orchestration of cognitive activities and text quality varied between tasks. Results indicated that this was indeed the case: the relation between each activity and text quality varied both over time and between tasks. Thus variations in process execution appear to be related to variations in text quality, although this depends to some extent on the moment at which activities are carried out. In other words, the moment at which an activity occurs (i.e., the timing of the variation) is of utmost importance. This should receive more attention in educational settings. Earlier research suggests that: *"it is advantageous to explicitly and systematically teach adolescents the processes and strategies involved in writing"* (Graham & Perin, 2007, p. 467). Informing students of the moments at which each activity appears to be most fruitful, could prove to be an effective addition. To start with the obvious, our results confirm that students should be encouraged to read the assignment carefully at the start of the writing process and to reflect on the information it contains. In addition, they should attempt to generate ideas related to the topic, before starting to write. Delaying formulating or text production until you have a clear picture in mind of what the task entails and what you think of it is likely to have a positive influence on the quality of the resulting text. Planning, however, is extremely task-dependent. The extent to which students need to plan what they will write and when it is advisable to plan, depends to a great extent on the task being carried out. This means that students should not be forced to plan their complete text at the start of the writing process. In some cases, delaying Planning until later stages of the writing process might be more advisable.

Finally, we would like to point out that it may seem tempting to infer causal relations between the occurrence of cognitive activities and text quality based on the results of this study, but we cannot do so. The results for Planning for example clearly indicate that it is not always easy to infer such relations and that we cannot

do so based on the data presented here (see also Hayes & Gradwohl Nash, 1996). Nonetheless, the fact that this study included multiple tasks per writer, makes this a first step on the road towards understanding the complex relationship between the writing process (the occurrence of cognitive activities) and its product (the quality of the resulting text). However, whether or not the process-product relations presented here are causal relations has yet to be determined. Empirical research aimed at answering that question is currently in progress.

APPENDIX A: Example of an assignment

Legalising soft drugs
The NSU, the National Student Union, is organising a national essay contest, especially for students. You're also taking part. You absolutely want to win. The winning essay will be printed in all the university newspapers, including the U-blad. The U-blad is read by students and employees of the university.

The subject of the essay has already been decided and was described in the U-blad as follows:
Soft drugs, such as marihuana and hash, are officially illegal in Holland, but are actually tolerated in practice. The legalisation of soft drugs has been the topic of fierce debate over the past few years. Some feel that the toleration policy should be abolished, whereas others feel that all drugs should be legalised, including hard drugs. The NSU would like to know how students feel about this topic. That's why they want to pay attention to this subject in a special edition of the U-blad. We want to hear from students what they think. Decide what you think and send us your response!

Assignment:
Write an essay in which you give your opinion on the question:
"Soft drugs: should they be legalised or prohibited completely?"

The essay has to meet the following requirements, set by the Jury:

1. Your essay must be (about) half a page in length.
2. You must do your best to convince your readers, readers of university magazines, of your opinion.
3. You must give arguments to support your opinion.
4. Your essay must be structured in a good and logical way.
5. Your essay must look well-cared-for (think of language use and spelling).
6. In your essay you must use at least two extracts from the 'References' (see next page). You must include these extracts in your essay in a meaningful way.

You have 30 minutes to complete this assignment.

Good luck!

References

The cabinet is strongly divided over the soft drugs policy. […] Alexander Pechtold, Secretary of State, pleads for a complete legalisation of soft drugs in the whole of Europe. Until that is the case, he feels that the sale of soft drugs should be moved to the outskirts of towns. […] Piet Hein Donner, Minister of Justice, opposes Pechtold, and doesn't feel anything for the proposals made by mayors in Southern Limburg to legalise weed cultivation. According to the mayors, that would decriminalize the weed cultivation, but Donner disputes that. "If that happens, Southern Limburg will become the drug state of Europe", according to the Minister of Justice.
Adapted from: www.nu.nl, April 25th 2005.

Once again our ambiguous soft drugs policy, which permits the sale of hash and weed, but prohibits its transportation and sale, has been ridiculed both in and outside the Lower House. The absurdity of the policy is undeniable, but those who desire consistency and an unambiguous policy in either direction, shouldn't harbour the illusion of having solved even half the problem by doing so.
Source: www.trouw.nl, April 28th 2005.

The "stepping stone" theory, states that people will switch from one drug to the next and finally end up in the hard drug scene. "It starts with a joint in the hand and finishes with a needle in the arm," as it were…. The most important thing we can say about this theory is that it is flawed. It has been scientifically proven, several times over, that soft drugs won't make you desire hard drugs. No one has become addicted to alcohol after drinking coffee. However both are drugs, one hard and the other soft….
Adapted from: Stefan Meyfroodt, www.pregonet.be, August 17th 2003.

According to a study carried out by Maurice de Hond, nearly half of all Dutch people (49 percent) feels that soft drugs should be legalised. A third desires a stronger policy regarding soft drugs. A clear majority (63 percent) is against weed cultivation in people's own homes. Source: www.nu.nl, April 27th 2005.

Once youngsters have started doing soft drugs, the switch to hard drugs is only a small step, according to those who oppose any form of legalisation. This "stepping stone" theory, is the main reason why many countries oppose drugs of any kind. The experiences in Holland however appear to show that the switch isn't automatic. "By far the largest part of cannabis users [stop] using before they're thirty, without ever having used hard drugs at all," according to the Ministry of Justice in a report called "The path to the backdoor". Source: www.nrc.nl, March 8th 2001.

"What is the debate actually about? Firstly, you can deny that soft drugs do actually have several pleasant and seductive properties. I don't have a problem with that. Secondly, you can't become physically addicted to soft drugs. Thirdly, one is hardly ever confronted with the alarming signals of acute intoxication. If one takes these three aspects together, then it is indeed clear that one should plead for legalisation, but at the same time one forgets the problems of psychological addiction and one underestimates several objective toxic effects."
Source: Psychiatrist Stan Ansoms, Veto nr 24, volume 22, March 11th 1996.

APPENDIX B: Analytical rating form

Read each essay carefully and then assign a score on each of the points below. Circle the score of your choice. The scale runs from 1 (does not meet the set criterion at all) to 5 (meets the set criterion in every way).

Part	**Score**
1. Structure	
1.1 Title The essay has a title which clearly corresponds with the content of the essay.	1 2 3 4 5
1.2 Structure The essay contains a clear division in: introduction, argumentation and conclusion.	1 2 3 4 5
1.3 Lay-out The essay is well-organized. There is a clear division in paragraphs. Paragraphs are divided by: a blank line, indenting or are started on a new line.	1 2 3 4 5
1.4 Subtopic Every paragraph has its own (sub)topic.	1 2 3 4 5
1.5 Relationships between Paragraphs There is a clear 'train of thought' between paragraphs: it is easy to determine coherence relations between paragraphs in the text.	1 2 3 4 5
1.6 Continuity Information which belongs together is presented together.	1 2 3 4 5
2. Content	
2.1 Introduction The thesis statement is presented in the introduction and (possibly) the writer's own opinion on the thesis statement is also provided.	1 2 3 4 5
2.2 Persuasion It is clear what the writer wishes to persuade the reader to believe: a choice for or against the thesis statement.	1 2 3 4 5
2.3 References The essay contains at least two quotes from the sources, which are incorporated in the essay in a meaningful way. For example, they support the argumentation or are used as an example in the introduction.	1 2 3 4 5
2.4 Refering (quoting from the sources) Quotes from the sources are correctly marked in the essay. Literal quotes (between inverted commas) and paraphrases both include references to the sources.	1 2 3 4 5

Part	Score
2.5 Reader Orientation The essay is easy to comprehend for readers who have not read the assignment. The writer does not, for example, refer to the assignment or to the experimental setting in the essay.	1 2 3 4 5
2.6 Reader Involvement The writer tries to engage the reader's interest by including real-life, common-place examples in the essay.	1 2 3 4 5
3. Argumentation	
3.1 Support The argumentation consists of multiple arguments, which support the writer's opinion.	1 2 3 4 5
3.2 Relevance The argumentation does not contain too much superfluous information, i.e., information which does not help support the writer's opinion.	1 2 3 4 5
3.3 Argumentation The arguments are clearly recognizable as such; for instance by the use of constructions such as *"therefore I do(not) think that ...", "I think ...", "I do (not) agree with ..."* etc.	1 2 3 4 5
3.4 Referential and Coherence relations The referential and coherence relations are clear if they are implicit, or else are marked explicitly. Examples of markings include: *therefore, so, thus, because, firstly, secondly, thirdly, subsequently,* etc.)	1 2 3 4 5
4. Conclusion	
4.1 Conclusion The essay contains a clear conclusion, which corresponds with the rest of the text and which indicates the writer's own opinion.	1 2 3 4 5
4.2 Closing It is clear that the essay is finished, for example by a brief summary of the content of the essay or by a closing statement. There are no loose ends left.	1 2 3 4 5

APPENDIX C: Structural framework for holistic rating

- For each essay, please determine to what extent it meets the requirements described in the framework and how much better or worse each essay is than the example essay on that topic. Assign each essay a score accordingly.
- The example essay is worth 100 points. If an essay is twice as good as the example essay, it is worth 200 points. If it is twice as poor, it is worth 50 points. All scores are possible. For example, an essay could be two-and-a-half times better (250 points) or worse (40 points) or three times better (300 points) or worse (30 points).

Example of Essay Structure

I <u>Introduction</u>
The thesis statement is introduced.
Optional: The writer states his opinion on the thesis statement.

A <u>Argumentation</u>
The writer is clearly for or against the thesis statement.
The arguments support the writer's opinion.

C <u>The writer's Conclusion</u>
The writer rounds off the essay with a clear conclusion.
The conclusion contains the writer's opinion (clearly for or against the thesis statement).

Clarification:

1. Compulsory order
I – A – C

2. Possible order variations
- Only arguments for or against the thesis statement;
- All arguments for the thesis statement followed by all the arguments against it;
- All arguments against the thesis statement followed by all the arguments for it;
- The arguments for and against the thesis statement are alternated;

3. Relationships
- If only arguments for or against the thesis statement are provided, then these must be linked in a clear additive sequence (e.g., *firstly, secondly, thirdly, furthermore, in addition,* etc.);
- If arguments are provided pair wise (*for, against, for, against*) then these must be presented as clearly opposing arguments (e.g., *but, pro-con, furthermore*, etc.)
- If both arguments for and against the thesis statement are provided, then the writer's own opinion (for or against) must become clear in the conclusion;
- There are clear/logical relationships present between the paragraphs;
- The relationships between the sentences within a paragraph are logical.

APPENDIX D: Examples of essays

Participant 18, task H: good quality essay (z-score: 1.04; rank: 13th of 80)

Against a ban on soft drugs

Soft drugs are tolerated in the Netherlands, which means that the sale of soft drugs is permitted. A long-standing debate is being carried out on this issue, with the main question being whether soft drugs should be legalised completely in the Netherlands or should be prohibited completely. Because soft drugs aren't as harmful as hard drugs, it is not necessary to prohibit them. However, complete legalisation could be a step too far.

First, it is important to state that soft drugs such as hash and weed are not actually harmful for people. They are not addictive in the same way that hard drugs are and proportionally do less harm. Besides, the so called "stepping stone"-theory, which states that soft drug use automatically leads to a craving for hard drugs, is not true.

Second, a ban on soft drugs would only lead to more drug related crimes. When its use is no longer legal, people will search for ways to keep using it illegally. In my opinion this is one of the advantages of the tolerance policy that the Netherlands currently has: there is no reason to use soft drugs through illegal means.

Moreover, research by Maurice de Hond indicated that 49 percent of the Dutch population thinks that soft drugs should be legalised. This is a significant number of people, and also clearly indicates that there is likely to be a lot of resistance, if soft drugs were to become prohibited.

A ban on soft drugs has more drawbacks than advantages. Because soft drugs are relatively harmless, the current policy prevents extra crime, and many Dutch people are pro legalisation, a ban would probably not work. Personally, I do not see the point in complete legalisation; the tolerance policy also appears to have enough advantages.

Participant 8, Task H: average quality essay (z-score .36; rank: 37th of 80)

Soft drugs aren't allowed in the Netherlands, but actually they are. This is the Netherlands' weak toleration policy. Soft drugs do not have fatal consequences. Also, it can't just be prevented. A country should have a clear policy regarding drugs. That is why I think that the Netherlands should legalise soft drugs.

First, scientifc research has proven that soft drugs do not have addictive effects and that they do not make you desire hard drugs. If people state that it should be forbidden, because it is addictive then they should also ban smoking, which does have addictive properties for that matter.

Furthermore, a country should have a clear drugs policy. The Dutch toleration policy is hypocritical. It is weak to forbid something, but to allow it because you can't take action against it. This makes the police seem a pointless organization. It is actually not possible to take action against it. People keep dealing and using it. It would be good for the Netherlands if she were to take hard action, but because I don't actually see the Netherlands doing so, legalisation is the best solution.

The legalisation of soft drugs in the whole of europe is not possible and also not advisable. The Netherlands has had a toleration policy for years and for other countries which have a strict drugs policy, legalisation would be too big a step.

In short, drug use cannot be prevented in a country such as the Netherlands. People will always continue using it. In addition, the fact that a country should have a clear drugs policy means that legalising soft drugs is the best solution.

APPENDIX E: Parameter estimates and standard errors (s.e.) in logits for all components of the models for all four cognitive activities

Fixed part of the model

Fixed part components	*Reading the assignment estimate (s.e.)*	*Planning estimate (s.e.)*
$\beta_0 * M^0_{ijk}$	-5.195 (.327)	-3.474 (0.103)
$\beta_1 * M^1_{ijk}$ [7]	-0.549 (.093)	-0.040 (0.020)
$\beta_2 * M^2_{ijk}$	.093 (.004)	---

Fixed part components	*Generating estimate (s.e.)*	*Formulating estimate (s.e.)*
$\beta_0 * M^0_{ijk}$	-3.037 (0.136)	-1.586 (0.056)
$\beta_1 * M^1_{ijk}$	-0.069 (<.001)	-0.202 (0.033)
$\beta_2 * M^2_{ijk}$	-0.012 (<.001)	0.029 (0.005)
$\beta_3 * M^3_{ijk}$	---	0.015 (0.001)
$\beta_4 * M^4_{ijk}$	---	-0.004 (<.001)
$\beta_5 * M^5_{ijk}$	---	0.000 (<0.0001)

[7] This value is centred on the mean (432) and divided by 100, which is merely a linear transformation for the sake of convenience; otherwise the estimates would be extremely small.

Random Part: Interwriter variance

	Reading the assignment estimate			*Planning estimate*		
Random part components	$S^2\ \beta_1$	$S^2\ \beta_2$	S^2*Text quality*	$S^2\ \beta_1$	$S^2\ \beta_2$	S^2*Text quality*
$S^2\ \beta_0$	1.531			0.155		
$S^2\ \beta_1$	0.424	0.120		-0.003	0.004	
S^2*Text quality**	-0.065	-0.031	0.626	-0.041	-0.003	0.626

* = *Standardized*

	Generating estimate			*Formulating estimate*		
Random part components	$S^2\ \beta_1$	$S^2\ \beta_2$	S^2*Text quality*	$S^2\ \beta_1$	$S^2\ \beta_2$	S^2*Text quality*
$S^2\ \beta_0$	0.370			0.047		
$S^2\ \beta_1$	0.028	0.014		0.015	0.017	
S^2*Text quality**	-0.102	-0.081	0.626	0.059	0.007	0.626

* = *Standardized*

Random Part: Intrawriter variance

	Reading the assignment estimate			*Planning estimate*		
Random part components	$S^2\ \beta_1$	$S^2\ \beta_2$	S^2*Text quality*	$S^2\ \beta_1$	$S^2\ \beta_2$	S^2*Text quality*
$S^2\ \beta_0$	2.030			0.140		
$S^2\ \beta_1$	0.557	.158		-0.009	0.006	
S^2*Text quality**	-.0001	-.006	0.132	-0.048	-0.022	0.132

* = *Standardized*

	Generating estimate			*Formulating estimate*		
Random part components	$S^2\ \beta_1$	$S^2\ \beta_2$	S^2*Text quality*	$S^2\ \beta_1$	$S^2\ \beta_2$	S^2*Text quality*
$S^2\ \beta_0$	0.226			0.019		
$S^2\ \beta_1$	0.036	0.022		0.002	0.003	
S^2*Text quality**	0.073	0.018	0.132	0.022	0.013	0.132

* = *Standardized*

CHAPTER 3

DIFFERENCES IN PROCESS AND PROCESS-PRODUCT RELATIONS IN L2 WRITING

Abstract [1]

The aim of this study is to examine whether writers vary how they write under influence of the changing task situation when writing in a second language (L2) and, if so, whether differences in the way they write are related to variations in text quality. Twenty participants, first year BA English majors, wrote four texts each in their L2 (English) under think-aloud conditions. The analysis focused on the occurrence of four cognitive activities: Reading the assignment and sources, Planning, Generating ideas and Formulating. Results indicate that, on average, the occurrence of each cognitive activity varies during the writing process. In addition, writers differ in the extent that they vary their process execution (i.e., the way they apply different cognitive activities) while writing. These differences, however, depend on the moment in the writing process, and on the activity being carried out. Furthermore, writers appear to vary the execution of the different activities between tasks, although results indicate that overall writers' behaviour is rather stable between tasks for most activities, at least in the beginning of the writing process. Finally, for all four activities, a correlation with text quality was calculated. Results indicate that this correlation varies during the writing process and also differs somewhat between tasks.

Keywords: Formulating; Generating ideas; L2 writing; Planning; Text quality;

In the past, writing process theories have often stressed that the task representation is an essential feature of the writing process (see for example, Carey & Flower, 1989; Flower & Hayes, 1980a). At the start of the writing process, writers form a representation of the task at hand. Subsequently, this representation changes as they write under influence of various factors, such as topic knowledge, reading the text-produced-so-far, rereading the assignment, or generating new ideas (Carey & Flower, 1989; Van der Hoeven, 1996a). As a result, they must continuously adapt the way they write (i.e., how they employ the cognitive activities needed during the writing process) to complement their revised task representation. As the task representation changes, certain cognitive activities might become more relevant

---

[1] Van Weijen, D., Van den Bergh, H., Rijlaarsdam, G., & Sanders, T. (in press). Differences in process and process-product relations in L2 writing. *ITL - International Journal of Applied Linguistics*.

than other activities, and thus occur more or less frequently at specific moments during the writing process (see Rijlaarsdam & Van den Bergh, 1996, p. 107). The important role the task representation plays during writing is indicated by the temporal distribution of such cognitive activities (i.e., the moment at which they occur) during the writing process. Research on first language (L1) writing has shown, for example, that an activity such as planning is much more likely to occur at the start of the writing process than near the end. Furthermore, the relations between the writing process and its product, the text, have been shown to depend on representations of the task, and thus depend on the moment at which an activity occurs during the writing process (see for example, Breetvelt, Van den Bergh, & Rijlaarsdam, 1994; Levy & Ransdell, 1995; Rijlaarsdam & Van den Bergh, 2006; Van den Bergh & Rijlaarsdam, 2007; Van der Hoeven, 1996a; see also Chapter 2). In other words, the moment at which activities are carried out is essential when assessing process-product relations in L1 writing: *"If the moment an activity is employed is left out of the analysis, hardly any relation can be found between cognitive activities on one hand and text quality on the other. Therefore, the moment a cognitive activity takes place is crucial"* (Rijlaarsdam & Van den Bergh, 2006, p. 42).

This does not hold only for Planning. Research on L1 writing found, among other things, that some activities, such as Reading the assignment and Goal Setting are more likely to occur at the start of the writing process, while others, such as Formulating and Revising are more likely to occur during later phases of the writing process (see Breetvelt, et al., 1994; Chapter 2). Furthermore, the temporal distribution of these activities was found to differ both between writers and between tasks (see Chapter 2). In addition, the correlation with text quality was shown to differ between activities and also varied over the writing process (see Breetvelt, et al., 1994) and between tasks as well (Chapter 2). For example, in L1, we found that Reading the assignment and Planning only correlated positively with text quality at the start of the writing process, while for Generating the correlation was positive in the beginning and the middle of the writing process (see Chapter 2). Formulating, on the other hand, correlated positively with text quality when it occurred after the start of the writing process.

Evidently, the role of task representation must be taken into account during writing process research. This is done by using the moment that an activity occurs as *"an indicator of the changing task situation, or perhaps to be more precise as an indicator of re-representations of the text produced so far."* (Van den Bergh & Rijlaarsdam, 1999, p. 103).

Even though the picture for L1 writing has become somewhat clearer over the years, less is known about L2 writing. In the past, research on L2 writing has focused on diverse topics such as: L2 writing instruction (e.g., O'Brien, 2004; Graham & Perin, 2007), comparisons of L1 versus L2 writing processes (e.g.,

Arndt, 1987; Beare & Bourdages, 2007; Chenoweth & Hayes, 2001; Roca de Larios, Manchón, & Murphy, 2006; Roca de Larios, Marín, & Murphy, 2001; Thorson, 2000), comparisons of different groups of L2 writers (e.g., Sasaki, 2002, 2004; Shi, 2004; see also Cumming, 2001, p. 5 for an overview), text analytical comparisons of L1 and L2 writing products (e.g., Hinkel, 2002), the use of L1 during L2 writing (e.g., Beare & Bourdages, 2007; Qi, 1998; Wang & Wen, 2002; Woodall, 2002), defining the construct of L2 writing (e.g., Di Gennaro, 2006), or on building models of L2 writing, either from scratch or based on existing L1 models (see for example, Cumming, 2001; Grabe, 2001; Roca de Larios, Murphy, & Marín, 2002; Silva, 1993; Wang & Wen, 2002). However, research on the relationship between the writing process (i.e., the cognitive activities which play a role during writing) and the quality of its resulting product (the text) in L2 has been somewhat scarce.

The few studies that did concentrate on process-product relations in L2, have either focused on the role of a single cognitive activity instead of the writing process as a whole (e.g., Jones & Tetroe, 1987; Broekkamp & Van den Bergh, 1996; Roca de Larios, et al., 2001; Stevenson, Schoonen, & De Glopper, 2006), on the effect of other process characteristics such as fluency or metacognitive knowledge on text quality (e.g., Schoonen, Van Gelderen, De Glopper, Hulstijn, Simis, Snellings, & Stevenson, 2003; Stevenson, 2005) or have concentrated on relating the writing process to specific product characteristics other than text quality, such as accuracy, syntactic complexity and fluency (e.g., Ellis & Yuan, 2004, 2005).

The only studies that we know of to date that compared multiple cognitive processes in L1 and L2 to text quality in both languages are Couzijn, Van den Bergh, & Rijlaarsdam (2002) and Uzawa (1996). Both studies were based on a single task per student per language, although Uzawa (1996) included an additional translating task per participant. But such a design makes generalizing on the basis of their results difficult, because their data provide no indications of variation between tasks. Furthermore, most comparison studies on L1 and L2 writing processes focused on quantitative differences between L1 and L2 (i.e., amount of planning or formulating, see Woodall, 2002) instead of on differences in relation to the changing task situation (i.e., differences in the moment at which activities occur). This means that we do not yet know whether writers also vary their behaviour under the influence of the changing task situation in L2 as they do in L1.

Research question and hypotheses

In their review of recent literature, Roca de Larios, et al. (2002) suggest that some of the limitations of earlier L2 studies could be solved by including the temporal nature of the writing process (a proxy variable for the changing task representation) as a variable in L2 writing research (Roca de Larios, et al., 2002, p. 44). That is why we will include time as a factor in this study, which leads us to the following main question: *Are changes in task representation reflected in the writing process in L2 writing?* In

other words, do writers vary their process execution (i.e., the way they apply different cognitive activities during the writing process) under influence of the changing task situation and, if so, are these variations in process execution related to variations in text quality?

In this study, the analysis of process execution focuses on four cognitive activities: *Reading the assignment, Planning, Generating ideas* and *Formulating*. We chose to include two conceptual activities, Planning and Generating, which are rather language independent processes, and two linguistic activities, Reading the assignment and Formulating, which are clearly language dependent processes (see also Levelt (1989)). These activities were also chosen because they each make an important contribution to the writing process, albeit in different ways. Reading the assignment is usually the first thing student-writers do when they are given a task by their teacher and thus it is one of the first possible factors that influences a writer's task representation (see Kennedy, 1985). Furthermore it is especially important in educational settings. Planning was included in the analysis because it is one of the most important elements in the writing process (see for instance Flower & Hayes, 1980a, 1981a; Breetvelt, et al., 1994; Levy & Ransdell, 1995; Van den Bergh & Rijlaarsdam, 2001) and because it appears to govern the writing process as a whole. We included Generating in the analysis because the presence of ideas is a prerequisite for text production (see also, Van den Bergh & Rijlaarsdam, 1999). Finally, we chose to include Formulating as it is the most important activity of all, in which words are formed into sentences and the text is produced. In addition, it is likely to vary strongly between and within writers during writing in L2.

First of all, the main hypothesis is that the temporal distribution of each cognitive activity (i.e., when it occurs during the writing process) will vary during the writing process in L2. In other words, each cognitive activity has specific moments when it is more likely to occur than at other moments during the writing process. Second, the extent to which the temporal distribution varies over time will vary between writers. For example, some writers will concentrate more on Reading the assignment in the beginning of the writing process, while others will do so at later stages. Third, the temporal distribution of each cognitive activity will also vary between tasks during writing in L2. Writers are likely to vary when they concentrate on each activity depending on the task they are carrying out. For example, they will not focus on Generating at more or less the same time during each task. Fourth, the correlation between the occurrence of each activity and text quality will vary during the writing process in L2. This means that each activity will only have a positive correlation with text quality at specific moments during the writing process. Finally, the correlation between each activity and text quality will vary between tasks in L2.

METHOD

Design

Twenty first-year English students (L1 Dutch) participated in the study. They were nearly all female (85%)[2] and their average age was 18 years. The study was carried out during their first semester at university, during which they were all enrolled in a first-year course on Academic Writing in English. They were paid a small amount for their participation in the study.

Each participant was asked to write four short argumentative essays in their L2 (English), while thinking aloud. The time allotted for each writing task was 30 minutes and the texts were produced on the computer, during four individual sessions. During each session audio and video recordings were made of the writing process.

At the start of the experiment, participants were trained to work under think-aloud conditions. This training was carried out individually by each participant during their first writing session and consisted of an algebra puzzle, a crossword style puzzle and a short writing assignment. The writing task entailed writing approximately five lines on a randomly assigned topic. There were ten topics in all and participants were asked to describe things such as their favourite music, film, book or what they did on their last birthday. To re-familiarize them with thinking aloud, participants were required to write a similar short text, on one of the other topics, at the start of each subsequent task.

The use of the think-aloud method in writing research is not unproblematic (see for example Ericsson, 1998; Roca, de Larios, et al., 2006). But Janssen, Van Waes, & Van den Bergh (1996) found that one of the potential problems, reactivity, does not play as large a role as might be expected. They found, for example, that the difference in amount of time spent writing or pausing between the think-aloud and silent conditions was relatively small, whereas the correlations between pauses in the two conditions were relatively high. Combining thinking aloud with a writing task was expected to require more cognitive effort, which was reflected in longer pause times. But there is no evidence that fundamental production processes proceeded in a qualitatively different way in one condition versus the other. So although thinking-aloud during writing causes some differences, both between and within conditions, these do not appear serious enough to warrant the dismissal of the entire research method (see also Ericsson & Simon, 1980; Hayes & Gradwohl Nash, 1996; Levy, Marek, & Lea, 1996; Manchón, Murphy, & Roca de Larios, 2005). Moreover, the use of the think-aloud method has provided valuable results in many studies (see for example, Beare & Bourdages, 2007; Breetvelt, et al., 1994; Chenoweth & Hayes, 2001; Jones & Tetroe, 1987;

[2] The male-female ratio of the participants in this study reflects the male-female ratio of the student population at the institute where the study was carried out.

Levy & Ransdell, 1995; Couzijn, et al., 2002; Roca de Larios, et al., 2006; Stevenson, 2005; Uzawa, 1996; Wang & Wen, 2002). Finally, the method was used in this study because, if we want to observe the role that cognitive activities play during the writing process, it still appears to be the best method currently available.

Assignment

Eight assignments were used in the study, which only differed in their topics. As mentioned above, participants each wrote essays on four of the eight topics: *Education in English*, *Having Children, Organ Donation, Life as a Student, Surveillance Cameras*, *Mobile Phone Use*, *Downloading Music*, and *Soft Drug Legislation*. The tasks were randomly divided over participants in a balanced design (i.e., each topic occurred just as often in first or last place).[3] In every task, participants were asked to produce an argumentative essay for a fictive essay contest organised by the local University Newspaper. The essay had to include the writer's own opinion as well as arguments for or against the given topic. The assignment contained information on the actual task (e.g., the topic, the goal and the target audience (peers and university employees)) as well as six concise sources on the topic. The sources were provided in order to give participants some background information on the topic, in case they were unfamiliar with it. Furthermore, they had to quote at least two of the sources in each essay.

Data

After the data collection was completed, we transcribed the audio and video recordings of each individual session to form think-aloud protocols (N = 79)[4]. The protocols were then segmented; each protocol segment contained only one cognitive activity. Whenever a writer's attention shifted from activity A to activity B (e.g., from generating to formulating) the segment ended and the new activity was placed in a separate segment (see Breuker, Elshout, Van Someren, & Wielinga, 1986).

After segmentation, the protocols were coded using a coding scheme largely based on Hayes and Flower (1980) and used frequently in earlier studies (see for example, Breetvelt, et al., 1994; Couzijn, et al., 2002; Van den Bergh, 1999; Van den Bergh & Rijlaarsdam, 1999; Van der Hoeven, 1996a; Chapter 2). The main cognitive activities included in the coding scheme are shown in Table 1. Research on L1 writing has shown that these eleven cognitive activities explained about 80% of the variance in text quality (see Breetvelt, et al., 1994). The protocols were coded by six different coders. Each coder coded 16 protocols, of which four overlapped with other coders, so that we could determine the interrater reliability, which was high (Cohen's kappa = 0.95).

---

[3] An example of an assignment can be found in Appendix A.

[4] One participant only wrote three essays instead of four due to technical difficulties.

Table 1: Overview of the eleven main categories included in the coding scheme

Cognitive activity	*Description and protocol example*
(Re)-Reading the assignment	Reading the assignment and/or the sources
Planning: Self-instructions	Instructions the participant gives himself regarding the next step in the writing process (e.g., *"I need to decide how to start"*)
Planning: Goal setting	Paraphrasing task demands in the assignment or setting new goals (e.g., *"This text is going to be about..."*)
Planning: Structuring	Selecting and ordering ideas or making a rough outline (e.g., *"First the pro's, then the con's and then my own opinion!"*)
Generating ideas	Paraphrasing ideas from the assignment or generating new ones: (e.g., *"Us humans we're very curious and we like to gossip!"*)
Metacomments	Reflections on the writing process as a whole or comments on the assignment and sources (e.g., *"I had the perfect sentence in my head"*)
Pausing	Silence or interjections (e.g., *um, ehhh*)
Formulating	Text production or typing the actual words on the computer.
Rereading own text	Rereading short or longer segments of the text produced so far
Evaluating own text	Evaluating the text produced so far at word, sentence or text level (e.g., *"No I'm not gonna say this"*)
Revising own text	Revising the text produced so far at word, sentence or text level (e.g., erases 'them' and types 'people')

In this study, we defined the cognitive activities included in the analysis as follows. Reading the assignment was defined as reading both the assignment or the sources aloud. Planning includes three sub-categories, namely self-instructions, goal setting and structuring. This combination was chosen because these three activities all appear to influence the other cognitive activities during the writing process (see for example Hayes & Gradwohl Nash, 1996, p. 33). This definition did not include Generating, which was seen as a different activity because it does not interact with other cognitive activities in the same way that planning does. Formulating, finally, was defined as the process of text production, during which words are formed into sentences and the text is typed on the computer (see also Van den Bergh & Rijlaarsdam, 2001).

In this study, the focus is on the moment at which each activity occurs, instead of on its overall frequency of occurrence. But to provide an overview of the

data, the average frequency of occurrence of each cognitive activity, its range between tasks and the average number of segments in each think-aloud protocol are provided in Table 2. These are all averaged over writers and tasks (*N* = 79) and are provided to give a general idea of the variation between and within writers.

Table 2: Average frequency of occurrence and range for each cognitive activity averaged over all texts and writers (N = 79)

		Range (number of segments)	
Cognitive activity	*Average number of segments (sd)*	*Min.*	*Max.*
Reading the assignment	37.20 (13.81)	12	88
Planning: Total	25.01 (19.59)	0	91
Planning: Self-instructions	23.13 (17.83)	0	84
Planning: Goal setting	0.76 (2.54)	0	19
Planning: Structuring	1.20 (2.21)	0	13
Generating ideas	41.51 (40.11)	1	215
Formulating	134.72 (39.69)	54	243
Writing process as a whole	875.73 (269.92)	390	1653

Table 2 shows that Formulating is, on average, the most frequent process of the four included in this study. It occurs, on average in 15.4% of segments, while Reading (4.2%), Planning (2.6%) and Generating (4.8%) occur far less frequently. In addition, self-instructions appear to be the most frequent form of Planning behaviour. Finally, the variation between writers and tasks appears largest for Formulating and Generating ideas, as these two activities have rather large standard deviations (see Table 2).

Text quality

Text quality was assessed in two different ways by the same jury of five judges: once analytically and once holistically. The 79 texts were first rated using an analytical rating form. This form had four main categories: *Structure, Content, Argumentation* and *Conclusion*. The starting point for its development was an earlier rating form (see Breetvelt, et al., 1994). The original categories in that form were modified so that the focus was on structure and argumentation (see Sanders & Schilperoord, 2006). The reason behind this revision was the idea that structure and argumentation are likely to be stable text features in texts written by the same writer, possibly even between languages. If this is the case, it could provide us with a tool for future studies on L1 and L2 writing. But that still has to be determined.

Table 3: Average text quality and range per topic in z-scores (N = 79)

Task	*Topic*	*N*	*Mean (sd)*	*Min.*	*Max.*
A	*Education in English: good idea or bad suggestion?*	9	-0.15 (1.31)	-1.96	1.72
B	*Having children, yes or no?*	10	0.17 (0.97)	-1.32	1.68
C	*Everybody should automatically be a donor, unless they explicitly register against it: good or bad idea?*	10	0.19 (1.09)	-1.49	1.82
D	*Life as a student: A hard life full of stress? Or isn't it that bad?*	10	0.31 (0.45)	-0.55	0.80
E	*Do surveillance cameras in inner city areas increase public security?*	10	0.15 (1.23)	-2.36	1.54
F	*Mobile phones: irritating or essential?*	10	-0.32 (1.34)	-3.08	1.53
G	*Downloading music for free: criminal or should be possible?*	10	-0.17 (0.53)	-1.37	0.66
H	*Soft drugs: should they be legalised or prohibited completely?*	10	0.18 (0.89)	-1.46	1.77
Total		79	0.05 (1)	-3.08	1.82

The second assessment was carried out two weeks later. For this holistic rating, each judge was given a benchmark essay (worth 100 points) for each topic. They had to decide to what extent each essay in the sample was better or worse than the benchmark essay and assign a score accordingly. The reliability between judges was acceptable for both methods (analytical: $\alpha = .93$ holistic: $\alpha = .83$) and the correlation between the two ratings was relatively high ($r = .69$; corrected for attenuation, $r = .79$). Consequently, a single general score for text quality was calculated, on an arbitrary (standardized) scale, based on the combined scores for the two different ratings. The overall average score and the average scores per topic are presented in Table 3 as z-scores, for ease of comparison. Although the mean scores are higher for some topics than others, the order in which the texts were written did not influence text quality scores.

Method of analysis

The data in this study are hierarchically structured on three levels: the participant level (n = 20), containing the task level (n = 79), which in turn contains the segment level (n = 69182). Therefore, we chose to perform a multilevel analysis, which enables us to do this hierarchy justice. Furthermore, multilevel analysis enables us to model the occurrence of each cognitive activity during the writing process. For this analysis, we chose to use polynomial models (see Breetvelt, et al., 1994); the occurrence of a specific activity is modelled as a function of when it occurs during the writing process. These polynomials (i.e., powers of the moment) are very flexible. Their shape depends on the number of powers in the model and their numerical values. For each activity, each segment Y is coded as 1 if a specific activity occurs and as 0 if it does not occur, at moment i (i = 1, 2, ... i_{jk}), during task j (j = 1, 2, 3, 4), for writer k (k = 1, 2, ... 20) during the writing process. Therefore, we can write the model as[5]:

$$Logit(Y_{ijk}) = \beta_{0\,jk} * M^{0}{}_{ijk} + \beta_{1\,jk} * M^{1}{}_{ijk} + \beta_{2\,jk} * M^{2}{}_{ijk} + \ldots \qquad (1)$$

The polynomial model can be adapted to include as many powers of the moment ($M^0, M^1, M^2, \ldots$) as are necessary to describe the data[6]. In general, no extra higher order coefficients are included in the model once a lower order coefficient lacks significance. Because the regression weights ($\beta_0, \beta_1, \ldots$) can vary both between tasks (j) and between writers (k), we can write them as:

$$\begin{aligned} \beta_{0\,jk} &= \beta_{000} + u_{0\,jk} + v_{00k} \\ \beta_{1\,jk} &= \beta_{100} + u_{1\,jk} + v_{10k} \\ &\ldots. \end{aligned} \qquad (2)$$

This means that the variance between tasks ($S^2\, u_{0jk}, S^2 u_{1jk}, \ldots$) and individuals ($S^2\, v_{00k}, S^2 v_{10k}, \ldots$) are estimated for each of the mean regression weights ($\beta_{000}, \beta_{100}, \ldots$). The first coefficient represents the extent to which the occurrence of a specific activity varies between tasks; the lower the value of this coefficient, the smaller the differences between tasks. The second coefficient represents the differences between individual writers. The extent to which each individual writer deviates from the general mean is represented by their residuals ($v_{00k}, v_{10k}, \ldots$). Furthermore,

---

[5] Please note: logit (P) = ln ($e^p/(1+ e^p)$)

[6] For the purpose of this analysis, segment number was used as a proxy variable for the changing task situation. This is not ideal, because it implies that all segments are equal in length. But it is a much better solution than dividing the writing process as a whole into three equal parts, as in earlier research (e.g., Breetvelt, et al., 1994).

the extent to which a specific task *j* deviates from the mean of a specific writer *k* is represented by the task and writer specific residuals (u_{0jk}, u_{1jk}, …). If we insert Equation 2 in Equation 1, we obtain the model to be estimated:

$$\begin{aligned} &Logit(Y_{ijk}) = M^{0}{}_{ijk}(\beta_{000} + u_{0\,jk} + v_{00k}) + \\ &M^{1}{}_{ijk}(\beta_{100} + u_{1\,jk} + v_{10k}) + \\ &M^{2}{}_{ijk}(\beta_{200} + u_{2\,jk} + v_{20k}) + \\ &.... \end{aligned} \tag{3}$$

The model, as shown in Equation 3, has a fixed and a random part. The fixed part (β_0, β_1, …) represents the mean change in occurrence of each activity over time, while the random part indicates the differences due to individuals and tasks. Both the variance between individuals and between tasks (random part) clearly depend on the moment (see Equation 3). More specifically, the variance between tasks at moment M, can be estimated as:

$$S^2_{task} = S^2_{u0\,jk} + 2 * S_{u0\,jk,u1\,jk} * M^1_{ijk} + S^2_{u1\,jk} * M^2_{ijk} + ... \tag{4}$$

We can use the same type of equation to note the variance between individuals, by replacing the residuals for tasks (u_{0jk}, u_{1jk} ...) with those that represent the deviations for individual *j* (v_{00k}, v_{10k}, ...).

A second model is used to estimate the relationship between the occurrence of each cognitive activity and text quality. Essentially, the model estimates the correlations between text quality and the residual score for task *j* (u_{0jk}, u_{1jk}, …) and the residuals for individuals (v_{00k}, v_{10k}, …). It is important to remember that these correlations depend on the value of M_{ijk} and its subsequent powers (see Van den Bergh & Rijlaarsdam, 1996).

In order to answer the main question, we must first determine the average temporal distribution for each cognitive activity in L2. This is the fixed part of the model. Next, we can determine whether there is any variation in process execution in L2 writing at all, and if so, whether there is mainly inter- or intrawriter variation or both. These questions pertain to the random part of the model. In the final step of the analysis, we will calculate the correlation between the occurrence of each cognitive activity and text quality at each moment in time. Then we can determine what the effect is of process variation on the relationship with text quality.

RESULTS

The parameter estimates of the model are presented in logits[7], a nonlinear transformation of proportions (see Equations 1 – 3 above). As this makes them hard to interpret, we have converted them to proportions, indicating how likely an activity is to occur at specific points during the writing process (cf. Fienberg, 1980). It is also important to note that the individual curves in some of the figures are not all the same length. Participants did not spend the same amount of time on each task, which means that the number of segments in each think-aloud-protocol varied, and thus the curves vary in length as well. Furthermore, a few participants took much more time to complete their essays than the rest of the group, which means that they had much larger numbers of segments (above 1200). This means that results for average curves after segment 1200 are based on observations of only a few writers and should thus be interpreted with caution.

Average temporal distribution and differences between writers

Figure 1 shows the average temporal distribution per activity, averaged over writers and tasks. In general, participants had the highest probability of occurrence for Formulating and Reading the assignment. Furthermore, it is also clear that the temporal distribution of each activity is different. Participants are most likely to focus on Reading the assignment at the start and end of the writing process, and are most likely to concentrate on Formulating in between. Finally, although the probabilities of occurrence for Planning and Generating are extremely low throughout the writing process, they are clearly most likely to occur in the beginning.

Figure 2 shows the average probability of occurrence per writer (averaged over four tasks) for each activity (n = 20). The figure shows that the patterns clearly differ between writers. For Reading the assignment, the differences between writers are large at the start of the writing process. Subsequently, they decrease, and almost disappear near the end of the writing process. Planning and Generating both follow a rather similar pattern. In both cases the probability of occurrence is generally low and differences between writers remain fairly constant throughout the writing process. For most writers the probability of occurrence appears to decrease slightly after the start of the writing process, although a few writers appear to do the reverse. Finally for Formulating, the variation between writers is rather large throughout the writing process. Formulating is most likely to occur in the middle of the writing process, after which most writers show a gradual decrease in their Formulating behaviour, with one or two exceptions.

[7] The parameter estimates in logits are provided in Appendix B.

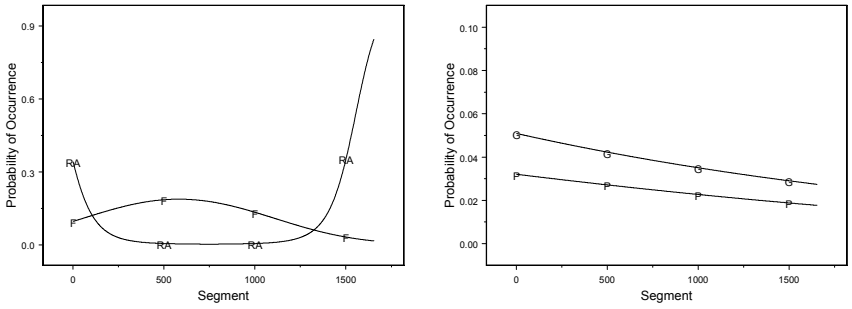

Figure 1: The mean change in the probability of occurrence of Reading the assignment (RA) and Formulating (F) (left) and Planning (P) and Generating (G) (right).

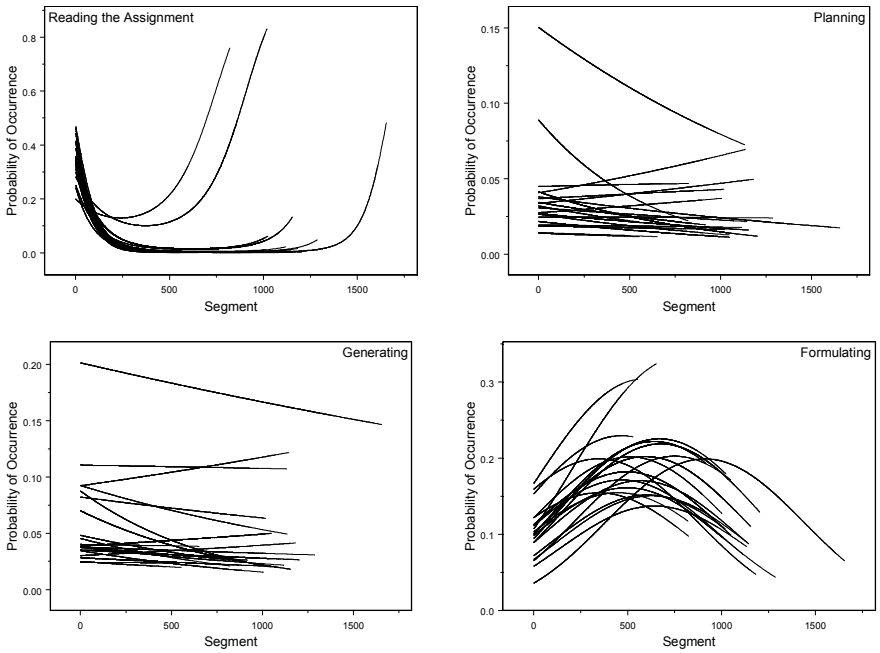

Figure 2: The average probability of occurrence for each cognitive activity per writer (n = 20).

Variation within writers

The next step is to examine the variation in the occurrence of each activity between tasks (within writers). In order to do so, we calculated the average probability of occurrence for each activity per writer and calculated the 80% confidence intervals. This indicates how much a writer's behaviour for a specific activity changes between tasks. If the 80% confidence intervals fall far from the mean then a writer's behaviour is likely to vary strongly between tasks. On the other hand, if they fall close to the mean then a writer's behaviour is likely to be rather stable between tasks. Figure 3 shows the average intrawriter variation (over all writers) for each activity (on the left) as well as the intrawriter variation for a specific writer (on the right). In the right hand panel of the figure, the letters on the curves (A – H) refer to the topic of the task (see Table 3).

The intrawriter variation for Reading the assignment is presented in Figure 3 (top row). On average, the pattern for Reading varies within writers. The differences between tasks are very large in the beginning and the end of the writing process, while there is almost no variation in the middle. The results for Writer 6 (see Figure 3, top row, right) further illustrate the within writer variation for Reading. For this writer the extent to which she is likely to read the assignment during different tasks clearly varies, although the curves for all tasks are very similar in shape. The differences are largest between Task F and the other three tasks, and smallest between Tasks A and B. So it seems that the intrawriter variation is largely due to her Reading behaviour in Task F.

In general, Planning is a rather infrequent activity. Figure 3 shows, however, that its probability of occurrence varies between tasks as well as over time (see Figure 3, second row, left). The within writer differences or task effects are generally largest near the end of the writing process. This is also illustrated by Writer 9 (see Figure 3, second row, right). Initially, the variation between tasks seems rather small for this writer, with only Task H differing markedly from the others. But subsequently the probability of occurrence for Tasks C and H decreases gradually over time, while for Task G it increases rather sharply as the writing process unfolds. If writers encounter problems while writing, they are likely to produce self-instructions (one of the activities we included as a type of planning), in an attempt to solve these problems.

The intrawriter variation for Generating appears to be fairly constant over time, on average, as it does not appear to be considerably larger at one specific point in the writing process compared to the rest. The results for Writer 1 (see Figure 3, third row, right), however, show that this does not hold for all writers in the study. For this writer, the variation between tasks appears somewhat larger near the end of the writing process than in the beginning. The intrawriter variation is strongly influenced in this case by Task H, for which she is clearly far more likely to generate new ideas throughout the writing process, than during her other three

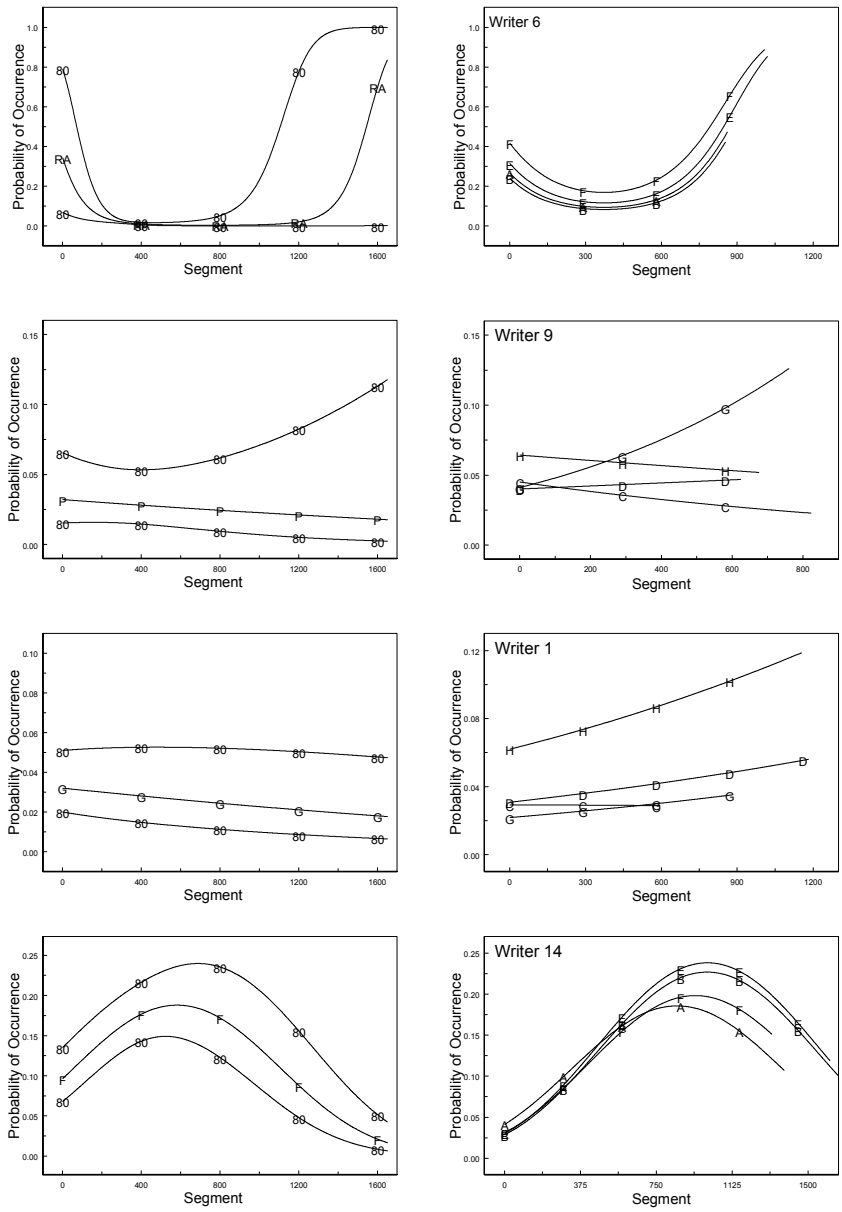

Figure 3: The intrawriter variation for each cognitive activity: Average probability of occurrence and 80% confidence intervals (left) Intrawriter variation for one specific writer (4 tasks) (right) (from top to bottom: Reading the assignment (RA), Planning (P), Generating (G) and Formulating (F)).

tasks. Although she does appear to generate more ideas for each task over time, except for Task C.

Finally, for Formulating, the intrawriter variation is highest, on average, in the middle of the writing process, after which it decreases slightly towards the end of the writing process. In other words, the probability of occurrence for Formulating varies within writers as well as over time, although the curves do resemble each other in shape. This is illustrated by the results for Writer 14 (see Figure 3, bottom row, right). This writer's between-task variation is rather small initially, although she is slightly more likely to formulate in the beginning during Task A than during her other three tasks. Subsequently, the differences between tasks increase slowly over time and become largest in the middle of the writing process, after which the probability of occurrence for all tasks decreases gradually over time.

Task-related variation

To determine the relative influence of individual or task-related variation on process execution, we calculated the intraclass correlations for each activity. The intraclass correlation is defined as the variation between individuals divided by the total variation (Snijders & Bosker, 1999). It indicates how much of the total variance is related to differences between writers. This coefficient can therefore also be used to indicate how different tasks affect the writing process[8]. The results of this analysis are presented in Table 4.

Table 4: The intraclass correlations indicating the proportion of between writer variation

	Proportion of variation between writers		
Cognitive activity	*Start*	*Middle*	*End*
Reading the assignment	.75	.62	.65
Planning	.68	.61	.45
Generating	.71	.70	.58
Formulating	.77	.33	.73

Table 4 shows that overall the proportion of variation due to individuals is much larger than the proportion of variation due to tasks. For Reading the assignment, the proportion of variation between individuals decreases over time, but is consistently larger than the variation due to tasks. For Planning, the proportion of

[8] 1 minus the intraclass correlation gives an indication of the proportion of variation due to tasks.

variation between individuals is largest at the start of the writing process after which it decreases gradually over time. Hence, for Planning, the influence of individuals is largest at the start of the writing process, while the influence of task conversely increases over time. Generating shows a similar picture to Planning, although the decrease in variation due to individuals is somewhat smaller over time. The proportion of variation due to individuals is largest in the beginning and subsequently decreases gradually over time. For Formulating, finally, the variation due to individuals is rather large and stable throughout the writing process.

All in all, the proportion of variation between individuals and between tasks both depend on the cognitive activity being carried out as well as on the moment at which they occur during the writing process. The variation between individuals is generally larger than the variation between tasks, particularly at the start of the writing process. Variation between tasks only seems to become important near the end of the writing process, although not for every activity. In other words, writers' behaviour appears rather stable between tasks for most activities, at least at the start of the writing process, which suggests that writers appear to start each task in a similar way.

Variation and text quality

The correlation between the occurrence of each activity and text quality was estimated in the last phase of the analysis. Because this correlation can vary between tasks, the average correlation was calculated as well as the 80% confidence intervals due to task differences. For each of the four activities we will first present the general correlation with text quality. Subsequently, we will show how the correlation between each activity and text quality varies between tasks.

Figure 4 shows that correlation between Reading the assignment and text quality (top left, solid line) varies from .30 to -.30 during the writing process. The correlation is only positive in the beginning, after which it becomes negative. This indicates that students who are likely to read the assignment at the start of the writing process are more likely to produce good quality texts. If students concentrate on Reading the assignment from the middle of the writing process onwards, however, this appears to be negatively related to the quality of their texts.

For Planning, the correlation with text quality (Figure 4, top right) varies from -.20 to .20 over the writing process. Initially, the correlation is negative (-.20), after which it gradually increases and becomes positive during the middle of the writing process. Furthermore, the correlation is strongest near the end of the writing process. Thus students who are likely to plan mainly in the beginning are likely to produce poorer quality texts than students who are more likely to plan from the middle of the writing process onwards.

The correlation between Generating and text quality is positive throughout the writing process (see Figure 4, bottom left, solid line). It is highest in

the beginning (.34), after which it gradually decreases over time to around .24 near the end of the writing process. Thus Generating ideas seems to be a useful thing to do throughout the writing process; its positive effect on text quality does not appear to be restricted to a specific phase during the writing process.

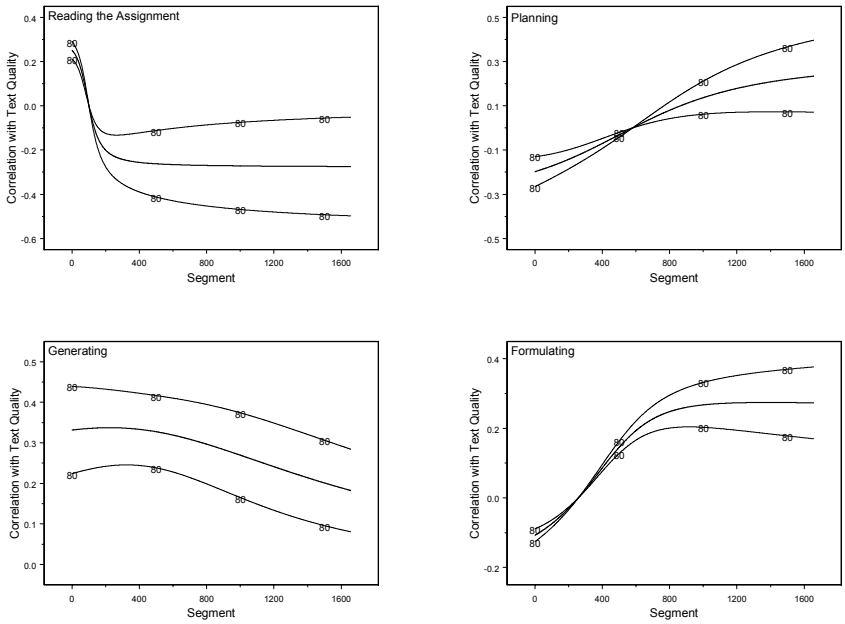

Figure 4: The correlation between Reading (top left), Planning (top right), Generating (bottom left), and Formulating (bottom right) and text quality: average (solid line) and 80% confidence intervals due to task (dotted lines).

For Formulating, the correlation with text quality varies from -.10 to .27 during the writing process. It is negative at first, but increases over time, becoming positive shortly after the start of the writing process. Subsequently, it remains positive over time, stabilizing around .27 after the middle of the writing process. It seems advisable, therefore, to focus on Formulating mainly from the middle of the writing process onwards. Students who are likely to devote most of their attention to Formulating after the start of the writing process are likely to have written better texts than students who are likely to formulate before then.

The correlation between each cognitive activity and text quality varies over time, but also between tasks. The extent to which it does so is indicated by the 80% confidence intervals in Figure 4. For Reading, the correlation with text quality varies somewhat in the beginning (range: .20 to .30), after which the differences between the confidence intervals decrease, disappear briefly, but then increase very rapidly and remain rather large from the middle of the writing process onwards

(range: -.05 to -.5). This indicates that the correlation varies strongly due to task at the end of the writing process, although it never has a positive effect on text quality. In other words, Reading the assignment only has a positive influence on text quality at the very start of the writing process. So, if writers are more likely to read the assignment at the start of the writing process during a specific task, compared to their other tasks, then that task is likely to have resulted in a better text. If we apply this to Writer 6 in Figure 3, then we would expect her to have produced her best text during task F, but this did not prove to be the case. This could be due to the fact that she had a very high probability of occurrence for Reading at the end of the writing process for that task, when the correlation with text quality is negative. That could have negated the positive effect of Reading the assignment in the beginning.

The correlation between Planning and text quality varies somewhat initially (range: -.12 to -.28), but the variation is largest near the end of the writing process (range: 0 - .40). Figure 4 shows that the correlation is only positive after the start of the writing process. So writers who are more likely to plan after the middle of the writing process than other writers are likely to have written better quality texts. But the correlation also varies strongly due to task, especially near the end of the writing process, when the range is largest. If writers encounter problems during the writing process, they might well use self-instructions, a type of planning, in order to get out of trouble. If that is the case, then planning is not always likely to result in good texts. This is the case for example, for Writer 9 (see Figure 3). Based on his relatively high probability of occurrence for planning during Task G, compared to his other three tasks, we would predict that this resulted in his highest quality text. But in fact the reverse is true, task G resulted in the poorest quality text this writer produced. Therefore, it is likely that he ran into trouble while writing and tried to get back on track by using self-instructions.

Although the correlation between Generating and text quality is consistently positive, Figure 4 shows that it varies strongly between tasks throughout the writing process. In other words, the correlation with text quality varies greatly due to task throughout the writing process. In general, students who are likely to generate more ideas for one task, above their own average, are likely to write higher quality texts. This is confirmed by the results for Writer 1 (see Figure 3). She wrote her best text during Task H, during which she was far more likely to generate ideas than during her other three tasks. By contrast, she produced her poorest quality text during Task C, when she hardly generated ideas at all.

For Formulating, finally, Figure 4 shows that the variation due to task is relatively small initially, but increases strongly over time and becomes largest near the end of the writing process (range: .18 to .38). This indicates that the correlation mainly varies due to task after the start of the writing process. So if writers are likely to formulate above their own average from the middle of the writing process

onwards, they are likely to have written higher quality texts. This also appears to hold for Writer 14 (see Figure 3) who produced her best text during Task B, when she is clearly far more likely to formulate than during tasks A and F.

CONCLUSION AND DISCUSSION

Process execution in L2

This study was set up to determine whether changes in task representation are reflected in the writing process in L2 writing. Based on our results we can draw a number of conclusions. First, we have shown that, on average, the occurrence of each cognitive activity varies during the writing process. The average probability of occurrence is different for each activity; for Reading it was highest at the start and end of the writing process, while for Formulating it was highest in the middle of the writing process. For Planning and Generating it was rather infrequent and varied over time. So for all four activities the way each activity is carried out seems related to the changing task situation, as in L1 writing.

Second, the general pattern of occurrence of each cognitive activity varies between writers. This variation depends on the activity being carried out as well as on the moment in the writing process at which it occurs. For Reading the variation between writers is largest in the beginning of the writing process. For Planning and Generating the variation between writers is related to the behaviour of only a few writers, while for Formulating the variation is rather large throughout the writing process.

Third, results show that the pattern of occurrence of each activity also varies within writers, between tasks. In other words, writers appear to vary the execution of the different activities due to the task at hand. Moreover, these differences in patterns of occurrence vary over time. For Reading, for example, the variation due to task is largest at the start and end of the writing process, while for Planning it is largest near the end of the writing process. For Formulating the variation is slightly higher in the middle of the writing process, while for Generating it is more or less stable over time.

Finally, the overall variation due to individuals seems much larger than the variation due to tasks, which suggests that writers' behaviour appears rather stable between tasks for most activities, at least at the beginning of the writing process. Indeed, process execution seems surprisingly stable within writers in L2, far more so than in L1 (see Chapter 2). Thus we can conclude that process execution varies both between and within writers, although this varies under influence of the cognitive activity employed as well as on the phase during the writing process.

Process variation and text quality

For all four cognitive activities, the correlation with text quality varied both over time and under influence of the task at hand. For Reading the assignment it is positive in the beginning and subsequently becomes negative, while for Generating it is positive throughout the writing process. For Formulating and Planning on the other hand, it is negative in the beginning and then becomes positive from the middle of the writing process onwards. For Reading the assignment, Generating and Formulating, the correlations with text quality appear similar to those found in L1 (see Breetvelt, et al., 1994; Rijlaarsdam & Van den Bergh, 2006). But for Planning the pattern was not the same as in L1.

Overall, the aim of this study was to examine whether writers vary their process execution under influence of the changing task situation and, if so, whether variation in process execution is related to variation in text quality. Based on our results, we can conclude that variation in process execution appears to be related to text quality. For instance, if a writer is more likely to generate ideas for one task than for his other tasks, then it is likely that this will have a positive effect on text quality (see Writer 1, Figure 3). Thus, for each of the cognitive activities discussed, adapting to the changing task situation seems appropriate. However, it is generally unwarranted to make such predictions based on a single cognitive activity, as activities do not occur in isolation during the writing process. Moreover, writers appear to have a preference for a specific orchestration of cognitive activities during the writing process; they seldom deviate from their default approach to the writing task.

However, our analyses focused on a selection of four cognitive activities. Whether our results also hold for other cognitive activities has yet to be determined. In addition, it seems as if the role variation plays in L2 is similar to its role in L1 writing, at least in general terms. But a full comparison of process-product relations during writing in both languages still has to be carried out.

Our results indicated that the correlation between each cognitive activity and text quality varies under influence of the task at hand. So, variations in the way cognitive activities are carried out appear to be related to variations in text quality. However, the variation we found appeared mainly to be due to individual variation rather than variation due to task, which means that individuals' behaviour was rather stable between tasks. Whether their behaviour was fairly stable due to the fact that we gave them rather similar assignments or due to the fact that they were writing in L2 is as yet unclear.

Finally, a word of caution is needed. We cannot conclude that process-product relations are causal in nature, based on the results of this study, even though it may seem tempting do so. In the case of Planning, for instance, self-instructions could be a signal that a writer is experiencing problems, and thus result in a decrease in text quality, while goal setting and structuring do appear to have a

positive influence on text quality (see Chapter 5). So variations in process execution and text quality are clearly related in some way. But whether variations in the way writers read, plan or formulate actually cause an increase or decrease in text quality has not yet been determined. This will be assessed experimentally in future research.

APPENDIX A: Example of an assignment

Surveillance cameras in inner city areas
The NSU, the National Student Union, is organising a national essay contest, especially for students. You're also taking part. You absolutely want to win. The winning essay will be printed in all the university newspapers, including the U-blad. The U-blad is read by students and employees of the university.

The subject of the essay has already been decided and was described in the U-blad as follows:
Due to the increase in crimes and meaningless acts of violence, more and more cities are choosing to place surveillance cameras in inner city areas. Not everyone is pleased about this. Some feel safe knowing that someone is 'watching over' them, while others consider it an invasion of their privacy. Because students often spend their free time in inner city areas, the NSU is going to pay attention to this subject in a special edition of the U-blad. We want to hear from students what they think. Decide what you think and send us your response!

Assignment:
Write an essay in which you give your opinion on the question:
"Do surveillance cameras in inner city areas increase public security?"

The essay has to meet the following requirements, set by the Jury:
1. Your essay must be (about) half a page in length.
2. You must do your best to convince your readers, readers of university magazines, of your opinion.
3. You must give arguments to support your opinion.
4. Your essay must be structured in a good and logical way.
5. Your essay must look well-cared-for (think of language use and spelling).
6. In your essay you must use at least two extracts from the 'References' (see next page). You must include these extracts in your essay in a meaningful way.

You have 30 minutes to complete this assignment.

Good luck!

References

Surveillance cameras help prevent crime, but they also increase the chance of tracking down the culprits. […] The cameras are placed in such a way that perpetrators of crimes within the inner city area are almost always registered. […] Incidentally, despite the presence of surveillance cameras, the security and well-being of the general public remains everyone's concern. That's why we still say: "if you spot trouble, warn the police!"
Source: Maastricht County Council, www.maastricht.nl, 2004.

The evaluation report of the project in Ede already mentioned that surveillance cameras don't just take away 'feelings of unease and insecurity.' On the contrary, before the cameras were installed, 65% of people visiting the Museumplein never felt unsafe, whereas five months after installation the percentage had dropped to 57%. Scottish research has also shown that after a short decrease the 'feelings of insecurity' rise again.
Source: Erik Timmerman, Leeuwarder Courant, July 7th 2000.

Great Britain has become THE surveillance capital of Europe, without anyone noticing, says Barry Hugill, spokesman for the English civil rights group Liberty. Remarkably, the call for privacy is gradually being overshadowed by experts' warning against 'a false sense of security.' Ian Brown, researcher for Information Policy Research […]: "It is an illusion to think that cameras will provide security." […] One study showed that in areas with intensive camera surveillance crime rates dropped by 3 or 4 percent, whereas better street lighting can help reduce the number of incidents by up to 20 percent. The general public is usually less vigilant, as the number of cameras in the area increases.
Adapted from: Steven de Jong, www.politiek-digitaal.nl, August 30th 2004.

The crime rate on the Wallen and in the vicinity of the Nieuwendijk in the centre of Amsterdam has decreased since the implementation of camera surveillance in March. […] In the Nieuwendijkkwartier especially, satisfaction prevails all round. […] In the vicinity of the Wallen, people are generally positive, although attention is drawn to the unwanted relocation of problems to other areas and changes in the group of troublemaking drug dealers and junkies. To combat these effects, the council will install three extra cameras.
Adapted from: Centrum voor Criminaliteitspreventie en Veiligheid, www.ccv.nu, November 30th 2004.

Security cameras in and around the Noorderstation are functioning with difficulty. It is even questionable whether the cameras, installed last year, are actually working. […] Not all the cameras are permanently on-line […] police spokesman Ed Kraszewski reported last week. They are chiefly there as a preventative measure. He believes that the presence of cameras, will deter thieves and violent criminals from committing criminal acts, even if the cameras are not on-line.
Source: Cees Vellekoop, Dagblad van het Noorden, August 18th 2003.

Permanent camera surveillance in Sneek's inner city is pointless. […] The city council based her decision, amongst other things, on the experiences of other city councils with security cameras. These experiences taught them that not all cities have had positive results and that the costs, especially personnel costs, have been substantial.
Adapted from: Friesch Dagblad, April 6th 2004.

APPENDIX B: Estimates and standard errors (s.e.) in logits for all components of the models for all four cognitive activities[9]

Fixed part of the model

Fixed part components	*Reading the assignment estimate (s.e.)*	*Planning estimate (s.e.)*
$\beta_0 * M^0_{ijk}$	-5.207 (0.062)	-3.571 (0.138)
$\beta_1 * M^1_{ijk}$ [10]	-0.501 (.000)	-0.034 (0.021)
$\beta_2 * M^2_{ijk}$	0.096 (.000)	---

Fixed part components	*Generating estimate (s.e.)*	*Formulating estimate (s.e.)*
$\beta_0 * M^0_{ijk}$	-3.110 (0.155)	-1.487 (0.070)
$\beta_1 * M^1_{ijk}$ [10]	-0.038 (0.017)	0.047 (0.017)
$\beta_2 * M^2_{ijk}$	---	-0.023 (0.001)

[9] Remember: logit (P) = $e^P/(1+ e^P)$

[10] This value is centred on the mean and divided by 1000, which is merely a linear transformation for the sake of convenience.

Random part: Variance between individuals

	Reading the assignment estimate			*Planning estimate*		
Random part components	$S^2\ \beta_1$	$S^2\ \beta_2$	S^2*Text Quality*	$S^2\ \beta_1$	$S^2\ \beta_2$	S^2*Text Quality*
$S^2\ \beta_0$	1.694			0.318		
$S^2\ \beta_1$	0.414	0.104		0.001	0.005	
S^2*Text quality**	-0.289	-0.076	0.725	-0.019	0.019	0.723

* = *Standardized*

	Generating estimate			*Formulating estimate*		
Random part components	$S^2\ \beta_1$	$S^2\ \beta_2$	S^2*Text Quality*	$S^2\ \beta_1$	$S^2\ \beta_2$	S^2*Text Quality*
$S^2\ \beta_0$	0.430			0.086		
$S^2\ \beta_1$	0.002	0.003		0.002	0.005	
S^2*Text quality**	0.183	-0.003	0.725	0.033	0.015	0.724

* = *Standardized*

Random part: Variance between tasks

	Reading the assignment estimate			*Planning estimate*		
Random part components	$S^2\ \beta_1$	$S^2\ \beta_2$	S^2*Text Quality*	$S^2\ \beta_1$	$S^2\ \beta_2$	S^2*Text Quality*
$S^2\ \beta_0$	1.023			0.180		
$S^2\ \beta_1$	0.226	0.051		0.013	0.007	
S^2*Text quality**	0.015	0.010	0.006	0.077	0.011	0.006

* = *Standardized*

	Generating estimate			*Formulating estimate*		
Random part components	$S^2\ \beta_1$	$S^2\ \beta_2$	S^2*Text Quality*	$S^2\ \beta_1$	$S^2\ \beta_2$	S^2*Text Quality*
$S^2\ \beta_0$	0.147			0.024		
$S^2\ \beta_1$	0.003	0.004		0.002	0.002	
S^2*Text quality**	0.002	0.003	0.006	0.004	-0.005	0.006

* = *Standardized*

CHAPTER 4

THE EFFECT OF WRITING IN L1 VERSUS L2 ON PROCESS-PRODUCT RELATIONS

Abstract [1]

Writing a text involves a number of cognitive activities, such as planning, formulating and revising. Research on first language (L1) writing has shown that the moment at which these activities occur during the writing process appears to influence the quality of the text produced. This study examines whether the relationship between the writing process and text quality changes when writers produce a text in a second (L2) or foreign language. Therefore, the main question is: *Does the effect of cognitive activities on text quality change when writers write in L2 instead of in L1?* Twenty students each wrote four short argumentative essays in L1 (Dutch) and four in L2 (English) under think-aloud conditions. The analysis focused on four activities: *Reading the assignment, Planning, Generating ideas* and *Formulating*. Results indicate that, the relations between the orchestration of each activity and text quality vary far less between tasks in L2 than in L1. In other words, writers' behaviour seems to become more consistent between tasks when writing in L2. As a result, it is possible to predict to some extent how writers write in L1 based on their L2 behaviour.

Keywords: L1; L2; Process Variation; Process-product relations; Text quality

Previous research on writing has shown that, basically, writing in a first (L1) or second (L2) language are two rather similar activities (see for example, Roca de Larios, Murphy, & Marín, 2002). In both languages, the writing process can be characterized as a recursive process, during which a number of different cognitive activities, such as planning, formulating and revising, are employed to produce a text (Silva, 1993). But there are also important differences, such as the influence of working memory capacity on writing in both languages. Research on L1 writing has indicated that working memory capacity plays an important role in the execution of the writing process (see Kellogg, 1994, 2001). It has been shown, for example, that an increase in task-complexity results in cognitive overload, making it harder for writers to maintain control over the execution of the writing process as a whole and thus produce good quality texts (Kellogg, 1994, 2001). Research has also shown that writers often produce texts of much lower quality in L2 (or in a foreign

[1] Van Weijen, D., Van den Bergh, H., Rijlaarsdam, G., & Sanders, T. (2008). *The effect of writing in L1 versus L2 on process-product relations*. Manuscript submitted for publication.

language) than might be expected based on their L1 performance (see for example Silva, 1993; Van Weijen, Van den Bergh, Rijlaarsdam, & Sanders, 2005a). This could be due to the fact that writing in L2 creates an added handicap, particularly for non-bilingual FL or L2 writers, as writing is usually a far more complex task in L2 than in L1. Ransdell, Levy and Kellogg (2002) found that L1 writers adapt the way they write, by pausing more often, as task demands increase. Writers in their study seemed able to distribute available working memory resources in such a way that text quality was least affected (Ransdell, et al., 2002, p. 159). However, this strategy failed once working memory resources became limited, due to extra task demands (Ransdell, et al., 2002, p. 160). This suggests that the increased cognitive load on working memory capacity, due to writing in L2, might have a negative effect on process execution and thereby also a negative influence on the quality of the resulting text. If we can determine how writers' process execution is linked to the quality of the products (i.e., the texts) they produce, and how this varies in L1 and L2, we might, in future, be able to predict how successful writers will be when writing in L2, based on their approach to the writing process in L1.

But so far, there is little agreement, as yet, on the question of how or why writing in both languages differs (see Silva, 1993, for an overview). Jones and Tetroe (1987), for example, suggest that differences and similarities are influenced by the type of comparison that is made: a quantitative or a qualitative one. According to them, research which focuses on quantitative comparisons of writing in L1 and L2 usually concludes that the two are very different, while research which focuses on qualitative comparisons often concludes that the two are, in general, very similar (Jones & Tetroe, 1987, p. 39). On the basis of their own data, they concluded that second language writing: *"is not a different animal from first-language composing"* (Jones & Tetroe, 1987, p. 55). But an unequivocal answer to the question in which respects L1 and L2 writing are comparable or different has yet to be found.

Studies that compared L1 and L2 writing in the past have either compared the way the writing process is carried out or compared the quality of the texts produced, but rarely combined the two approaches. Studies comparing process execution or the role of specific cognitive activities during writing in L1 and L2 have reported a number of differences: on average, students plan less (Jones & Tetroe, 1987; Silva, 1993), write less (e.g., Chenoweth & Hayes, 2001; Thorson, 2000), and revise more in L2 or FL (e.g., Stevenson, Schoonen, & De Glopper, 2006; Thorson, 2000). Furthermore, although they spend roughly the same amount of time formulating in both languages (Roca de Larios, Marín, & Murphy, 2001), they spent relatively far more time on solving problems while formulating in L2 than on fluent formulating (Roca de Larios, Manchón, & Murphy, 2006). In a very thorough overview of L2 writing process research, Roca de Larios, et al. (2002) conclude that differences between L1 and L2 are usually related to the frequency

with which certain activities occur during the writing process. Furthermore, differences in the frequency of occurrence between languages appear related to *'the constraining effect'* of writing in L2 (Roca de Larios, et al., 2002, p. 35).

Text analytical comparisons of L1 and L2 writing, also reported differences: L2 texts usually contain more errors, are shorter and are (judged to be) lower in quality (Silva, 1993). Furthermore, research has shown that L2 writers often use far more connectives than L1 writers (see for example, Degand, Demol, Hardermann, & Perrez, 2006; Hinkel, 2002; Perrez, 2006). Hinkel (2002) carried out a detailed and extensive comparison of English academic texts written by native and non-native speakers. She concluded that non-native speakers clearly: *"over rely on simple phrase- and sentence-level conjunctions and exemplification"* (Hinkel, 2002, p. 159). Furthermore she concludes that even trained L2 writers appear to lack the syntactic and lexical knowledge they need to write academic texts of high quality (Hinkel, 2002, p. 160).

Finally, a small number of studies was carried out on the differences between process-product relations (i.e., the relation between the occurrence of cognitive activities and text quality) in L1 and L2. Such L1 – L2 comparison studies usually focused on the role of a single cognitive activity, such as planning (Jones & Tetroe, 1987), or revising (Stevenson, et al., 2006) on text quality or on the effect of specific process characteristics such as metacognitive and linguistic knowledge or fluency on text quality (e.g., Schoonen, Van Gelderen, De Glopper, Hulstijn, Simis, Snellings, & Stevenson, 2003, Stevenson, 2005). As far as we know, there have only been two studies, thus far, which have related the role of multiple cognitive processes in L1 and L2 to text quality in both languages (Couzijn, Van den Bergh, & Rijlaarsdam, 2002; Uzawa, 1996). However, it is hard to compare the two, because Couzijn, et al. (2002) had a relatively small number of participants, while Uzawa (1996) used a single task per participant per condition design. Such a single task design is problematic, in general, because research has shown that writers' behaviour can vary strongly between tasks (see Meuffels & Van den Bergh, 2005; Chapter 2).

In sum, research comparing writing in L1 and L2 has focused almost exclusively on differences in process execution or on differences in text characteristics, but not on differences in process-product relations in both languages. In other words, the differential effect that process execution has on product quality in both languages has yet to be studied. Moreover, comparing the results of all these different studies and making generalizations based on their results has proved difficult for a number of reasons. First of all, the studies on L1 versus L2 writing mentioned above (comparing process characteristics, product characteristics or both) differ in many ways. They differ, for example, in: the number of participants, their age, schooling, and L1 or L2, and the research method (e.g., the use of: concurrent or retrospective protocols; key-stroke logging;

observations etc). They also differ strongly in the number and type of tasks writers have to carry out.

Second, studies that focus only on the writing process or on the texts produced, seem, in general, to be less informative than studies that focus on both parts of the equation (i.e., the process and its product). Indeed: *"it may be ill-advised, and perhaps even impossible, to divorce processes and products from each other, either in teaching or in research"* (Arndt, 1987, p. 257; see also Roca de Larios, et al., 2001, p. 528).

Third, most studies on L1 - L2 writing were based on between group comparisons instead of within-writer comparisons. Several studies reported, for example, that writing strategies were generally rather consistent across languages (see for example, Arndt, 1987; Beare & Bourdages, 2007; Raimes, 1994; Roca de Larios, et al., 2002; Silva, 1993). However, these studies consisted of a single task per participant per condition design, such as described above, or were based on between group comparisons instead of on within writer comparisons (e.g., Raimes, 1994). This means that the writing behaviour and/or texts of a (homogenous) group of L1 writers (usually English) is compared to the behaviour and/or texts of a (heterogeneous) group of L2 writers (writing in English as L2). Several studies have indicated that comparing the processes and products of the same writers in both languages is *"a significantly more effective research design for analyzing L1 and L2 writing processes than comparing across groups"* (Thorson, 2000, p. 156; see also Roca de Larios et al., 2002, p. 31; Zimmermann, 2000, p. 79). Therefore, whether writing strategies really are consistent across languages *within writers* remains an issue to address in future research.

Fourth, writers continuously revise the task representation that they form at the start of the writing process as they write. Many studies, however, did not include task representations in their analyses. Only a few studies have included the factor time in their analyses, as a proxy variable for the role of changing task representations during the writing process (see for example, Breetvelt, Van den Bergh, & Rijlaarsdam, 1994; Rijlaarsdam & Van den Bergh, 2006; Chapter 3). This means that they examined when specific cognitive activities occur during the writing process and how the way in which these activities occur varies over time. Roca de Larios, et al. (2002) also suggest that including time as a factor in future writing process research is important (Roca de Larios, et al., 2002, p. 44).

Finally, studies on process-product relations have an additional problem to deal with. Research has shown that determining writing ability can be incredibly problematic (see for example, Coffman, 1966; Godshalk, Swineford, & Coffmann, 1966; Schoonen, 1991, 2005; Van den Bergh, De Glopper, & Schoonen, 1988; Wesdorp, 1974). Using essay scores to assess writing ability is particularly problematic, because text quality can vary strongly within writers, due to factors such as topic, genre, time on task, scoring procedures, the language being written in and disagreements between raters (Schoonen, 2005, p. 2; see also Coffman, 1966;

Wesdorp, 1974). This implies that a writer's skill cannot be assessed accurately based on a single text. To adequately determine writing skill, multiple texts by the same author must be assessed by at least three to five raters (Coffman, 1966; Wesdorp, 1974). Estimates of the number of texts that must be examined vary from four (Coffman, 1966) to twenty texts (Van den Bergh, et al., 1988), depending on the type of task, ranging from general essays to more specific functional texts. Determining writing skill in two different languages is even more difficult, because it is not possible, in general, to compare L1 and L2 text quality directly (see for example, Stevenson, 2005; Stevenson, et al., 2006). The scores are usually not on exactly the same scale and research has shown that texts in different languages are usually not evaluated in a comparable manner (see Roca de Larios, et al*.,* 2002, for an overview).

Research question and hypotheses

All-in-all, it seems clear that the influence of language (L1 or L2), an important task feature, on process-product relations has yet to be analyzed in greater detail. Therefore, the aim of this study is to answer the following main question: *Does the effect of cognitive activities on text quality change when writers write in L2 instead of in L1?* We will first examine how process execution influences text quality in both languages, and then compare the results for L1 and L2, to determine how process-product relations in these two languages differ. In addition, we will examine to what extent writers' approach to writing in L1 can predict how successful they will be when writing in L2. The current study was designed to avoid as many of the limitations discussed above as possible. The study thus consists of an intrawriter comparison of L1 and L2 writing, with multiple tasks per writer per condition and includes an alternative method for comparing text quality scores across languages.

The present study will focus on the process-product relations of four cognitive activities in L1 and L2: *Reading the assignment, Planning, Generating ideas* and *Formulating*. These four activities were chosen for two main reasons. First, including two linguistic, language dependent processes (Reading the assignment and Formulating) and two conceptual, or fairly language independent, processes (Planning and Generating), in the analysis, might provide interesting results when comparing writing in L1 and L2 (see Table 1 for examples; see also Levelt, 1989; Stevenson, 2005). Second, these activities were also selected for their own contribution to the writing process. Reading the assignment was selected because it is the first input writers usually have for their task representations (see Kennedy, 1985). In addition, it plays an important role in educational settings, as the assignment is usually one of the first things writers read when given a writing task by their teacher. We included Formulating because it is the central component of the writing process, during which the ideas which are generated are put into words and the actual text is produced. We also included Formulating because, as a

language dependent activity, its orchestration might vary strongly between languages. Planning consists, in our view, of remarks concerning goal setting, structuring and self-instructions. It was included because it is one of the core components of the writing process (see for example, Breetvelt, et al., 1994; Flower & Hayes, 1980a; Levy & Ransdell, 1995; Van den Bergh & Rijlaarsdam, 2001) and it seems to exert influence over the writing process as a whole. Finally, Generating ideas was included in this analysis because the presence or absence of ideas to write about can have a large influence on the way the writing process unfolds (see Van den Bergh & Rijlaarsdam, 1999).

To answer the main question, we must determine the process-product relations in L1 and L2 for each cognitive activity, and then compare the relations in the two languages to determine how they differ. Process-product relations are analyzed to determine how the way in which each cognitive activity is carried out (i.e., the process) affects the quality of the text produced (i.e., the product). The term *orchestration of cognitive activities* is used in this study to describe the temporal distribution of cognitive activities over the writing process (Alamargot & Chanquoy, 2001; Braaksma, Rijlaarsdam, Van den Bergh, & Van Hout-Wolters, 2004, p. 2; Graham & Harris, 2000; McCutchen, 2000; McCutchen, Covill, Hoyne, & Mildes, 1994). We will first examine the orchestration of each cognitive activity (i.e., when it occurs during the writing process) and then examine how changes in the orchestration of each activity between languages affect the quality of the resulting text.

Our first hypothesis is that the orchestration of each cognitive activity will change during writing in L2, which means that the moment at which each activity is most likely to occur will differ between languages. More specifically, we predict that the orchestration of the linguistic activities (Reading & Formulating) is more likely to change, as a result of writing in L2, than the orchestration of the conceptual activities (Planning & Generating). The linguistic activities are language dependent, and will therefore be influenced more strongly by the language in which they are carried out, than the more language independent activities, such as Planning and Generating. Second, we predict that the extent to which writers vary the orchestration of activities between tasks will be greater in L1 than L2. This means that writers' behaviour is likely to be more consistent between tasks in L2 than in L1. The increased cognitive complexity of writing in L2 is likely to increase the writer's workload when working in L2, leaving less room in working memory for adapting to different writing tasks than during L1 writing. Thus resulting in fewer differences between tasks in L2 than in L1. Third, because there are changes in the way activities are orchestrated in L2, process-product relations (i.e., the effect of process execution on product quality) are likely to change as well. Therefore, we predict that the moment at which each cognitive activity is positively related to text quality will be different in L1 and L2.

METHOD

Design

The participants were twenty first-year English students (L1 Dutch), nearly all female (85%), and their average age was 18. They participated in the experiment during their first semester at university. At the time, they were all enrolled in a first-year course on Academic Writing in English and all received a small financial reward for their participation in the study. Please note that the high percentage of female participants is in line with the relatively high number of female students currently enrolled in the English Department.

Each participant wrote four short argumentative essays in their L1 (Dutch) and four in their L2 (English), while thinking aloud. The writing time per task was 30 minutes. The texts were produced on the computer, during four individual sessions. Participants wrote two texts per session, during which audio and video recordings were made of each writing process. The L1 and L2 texts were distributed across sessions in order to avoid order effects.

At the start of the experiment, each participant was individually trained to work under think-aloud conditions. The training consisted of an algebra puzzle, a crossword style puzzle and a short writing task. They were asked to write roughly five lines on one of ten randomly assigned topics (e.g., favourite music, book, film, or where they went on their last holiday). Participants wrote another short text on one of the other topics before starting each subsequent assignment, to re-familiarize them with thinking aloud. The think-aloud method was used in this study, because, currently, it still seems the best possible way to observe the execution of cognitive activities during the writing process[2]. Although keystroke logging is an excellent non-obtrusive on-line tool for studying writing processes, we decided against using it in this study, because it cannot provide information on activities which do not occur on the computer, such as Reading the assignment, Planning and Generating ideas.

Assignment

The eight different assignments included in the study were identical in every way, except their topic (see Appendix A). In every task, participants were asked to produce an argumentative essay for a fictive essay contest organised by the local University Newspaper. The essay had to include the writer's own opinion as well as arguments for or against the given topic. The assignment contained information on the actual task, such as the topic, the goal and the target audience (peers and university employees), as well as six concise sources on the topic. The sources were

[2] For a discussion on the pro's and con's of the method, see for example: Ericsson, 1998; Janssen, Van Waes, & Van den Bergh, 1996; Roca, de Larios, et al., 2006.

provided in order to give participants some background information on the topic, in case they were unfamiliar with it. Furthermore, they had to quote at least two of the sources in their essays. The following topics were included in the study: *Education in English*, *Having Children, Organ Donation, Life as a Student, Surveillance Cameras*, *Mobile Phone Use*, *Downloading Music*, and *Soft Drug Legislation*. To reduce the risk of task-order or language effects, we divided the topics across writers and between languages in such a way that every participant wrote about each topic in either L1 or L2 and that no topic occurred more often in L1 or L2 or in first or last place than the others.

Data

The audio and video recordings of each writing process were combined and transcribed to form think-aloud protocols (L1: $N = 80$, L2: $N = 79$)[3]. Next, the protocols were segmented, so that each protocol segment contained a single occurrence of a cognitive activity. Whenever a writer switched from activity A to activity B (e.g., from Planning to Formulating) the new activity was placed in a new segment (see Breuker, Elshout, Van Someren, & Wielinga, 1986). Subsequently, the segmented protocols were coded using a coding scheme that was loosely based on Hayes and Flower (1980) and used frequently in earlier research (see for example, Breetvelt, et al., 1994; Couzijn, et al., 2002; Van den Bergh, 1999; Van den Bergh & Rijlaarsdam, 1999; Van der Hoeven, 1996a; Chapter 2). The coding scheme consists of eleven main cognitive activities (see Table 1). The L1 and L2 protocols were coded by two separate groups of six coders. The interrater reliability between coders was relatively high (Cohen's kappa: L1 = .84, L2 = .95).

The four cognitive activities included in the analysis were defined as follows. Reading the assignment included any occurrence of reading the assignment or sources out loud. Formulating was defined as the process of text production, which includes sentence formation and typing the text on the computer. Planning was defined as all self-instructions, and all comments related to goal setting and structuring, because these three activities all appear to guide the other activities during writing (see also Hayes & Gradwohl Nash, 1996, p. 33). Generating was defined as a separate activity, because it deals primarily with the content of the text and does not seem to interact with the other activities in the same way as Planning.

This study will not focus on the frequency with which activities occur, but on the moment at which they occur during the writing process. But frequencies do provide some background information on the data. Therefore, to provide an overview of the L1 and L2 data, the average frequencies of occurrence for each activity, averaged over writers and tasks, and the average number of segments per think-aloud protocol for both languages, are presented in Table 2.

---

[3] One participant wrote three essays in L2 instead of four due to technical difficulties.

Table 1: Overview of the eleven main categories included in the coding scheme

Cognitive activity	*Description and protocol example*
(Re)-Reading the assignment	Reading the assignment and/or the sources
Planning: Self-instructions	Instructions the participant gives himself regarding the next step in the writing process (e.g., *"I need to decide how to start"*)
Planning: Goal setting	Paraphrasing task demands in the assignment or setting new goals (e.g., *"This text is going to be about..."*)
Planning: Structuring	Selecting and ordering ideas or making a rough outline (e.g., *"First the pro's, then the con's and then my own opinion!"*)
Generating ideas	Paraphrasing ideas from the assignment or generating new ones: (e.g., *"Us humans we're very curious and we like to gossip!"*)
Metacomments	Reflections on the writing process as a whole or comments on the assignment and sources (e.g., *"I had the perfect sentence in my head"*)
Pausing	Silence or interjections (e.g., *um, ehhh*)
Formulating	Text production or typing the actual words on the computer.
Rereading own text	Rereading short or longer segments of the text produced so far
Evaluating own text	Evaluating the text produced so far at word, sentence or text level (e.g., *"No I'm not gonna say this"*)
Revising own text	Revising the text produced so far at word, sentence or text level (e.g., erases 'them' and types 'people')

Of the four activities included in the analysis, Formulating is the most frequent activity, on average, in both languages (L1: 17.2%; L2: 15.4%; see Table 2). Furthermore, Planning occurs more frequently, on average, in L1 (L1: 5.3%; L2: 2.6%), while Generating occurs more frequently, on average, in L2 (L1: 3.1%; L2: 4.8%). The differences in Reading the assignment between the two languages are small (L1: 3.7%; L2: 4.2%). Finally, a comparison of the three activities that are included in the planning category reveals that self-instructions occur most frequently in both languages, and more frequently in L2 than in L1 (L1: 81%; L2: 92%).

Table 2: Average frequencies of occurrence and range for each cognitive activity in L1(n = 80) and L2 (n =79) averaged over all texts and writers

	L1 Dutch			*L2 English*		
Cognitive activity	*Average no. of segments (sd)*	*Range (no. of segments)*		*Average no. of segments (sd)*	*Range (no. of segments)*	
		Min	*Max*		*Min*	*Max*
Reading the assignment	26.46 (11.28)	10	77	37.20 (13.81)	12	88
Formulating	123.9 (44.52)	45	234	134.72 (39.69)	54	243
Planning: Total	22.50 (16.92)	0	98	25.01 (19.59)	0	91
Planning: Self-instructions	18.21 (12.74)	0	62	23.13 (17.83)	0	84
Planning: Goal setting	0.88 (1.93)	0	13	0.76 (2.54)	0	19
Planning: Structuring	3.41 (5.36)	0	28	1.20 (2.21)	0	13
Generating ideas	38.05 (36.53)	0	187	41.51 (40.11)	1	215
Writing process as a whole	721.53 (290.41)	222	1529	875.73 (269.92)	390	1653

Text quality

Comparing the quality of texts produced by the same writer in two different languages, in this case Dutch and English, is difficult for a number of reasons. For example, certain errors, such as spelling mistakes and other surface level linguistic errors, occur in L2 writing that do not (usually) occur in L1 writing. That means that texts in L2 will almost certainly be rated less highly than texts written in L1, making it hard to compare texts in the two languages. Therefore a method was used in which the influence of spelling errors was reduced to a minimum.

This method was developed based on the idea that it should focus on structure and argumentation (see Sanders & Schilperoord, 2006), as these seemed likely to be stable features in texts written by the same writer in L1 and L2. If so, that could provide us with a clear link between the two languages, which might

help us compare text quality scores. Because it was unclear which method would work best, we decided to develop both an analytical rating scheme and a more holistic rating method. The analytical rating form consisted of four main categories: *Structure, Content, Argumentation,* and *Conclusion* (see Appendix B), and was based on a form used in earlier research (see Breetvelt et al., 1994). For the holistic rating judges were provided with a benchmark essay on each topic, worth 100 points, to help them with their rating. They had to determine how much better or worse they considered each text to be than the benchmark essay (see Blok, 1986). In addition, they were provided with an overview of how argumentative essays should be structured (see Appendix C).

The L1 texts (n = 80) were first rated analytically by a group of five raters (third year BA students), followed by a holistic rating by the same raters, two weeks later. The between rater reliability was satisfactory for both methods (analytical: α = .88, holistic: α = .82) and the correlation between the two ratings was rather high (r = .74, or .87 when corrected for attenuation). Therefore, the scores were combined to form a single general score for each text on a standardized scale.

The L2 texts (n = 79) were also rated twice, by a group of five raters (MA students), and again the analytical rating was carried out roughly two weeks before the holistic rating. The between rater reliability was satisfactory for both methods (analytical: α = .93 holistic: α = .83) and the correlation was relatively high (r = .69; corrected for attenuation, r =.79). So again, a single score was calculated for each text.

In an attempt to make the L1 and L2 text quality scores comparable, we translated the Dutch (L1) benchmark essays into English (L2) and included them in the L2 rating without informing the raters. They judged the quality of 79 L2 essays and the 8 L1 benchmark essays translated into English. If the English versions of the L1 benchmark essays received higher scores in L2 than most of the L2 essays, this could provide us with an indication of the average difference in quality between the L1 and L2 texts. The difference in scores between the L1 and L2 benchmark essays was calculated, by converting the L1 and L2 scores into Z-scores. However, the results indicated that for two tasks, the Z-scores for the translated L1 benchmark essays were much lower than the scores for the L2 benchmark essays. Thus we did not find a consistent difference in text quality between the two languages across tasks. This made it unadvisable to compare text quality scores for a specific writer across languages. In other words, we cannot say that writer 14 writes better in L1 than in L2 or vice versa. But we can compare process-product relations in L1 to those in L2, and thus determine whether or not the occurrence of a specific activity correlates with text quality in the same way in both languages, which is the main aim of this study.

Method of analysis

Previous research has shown that multilevel modelling is an appropriate technique for analysing hierarchically ordered data (see for example, Braaksma, 2002; Breetvelt et al., 1994; Leijten, 2007; Snijders & Bosker, 1999; Van den Bergh & Rijlaarsdam, 1996). In this case, the data are hierarchically structured on three levels (from top to bottom): participant (n = 20), task (n = 159 (20 participants * 4 tasks * 2 languages)) and segment (n = 127070). So the protocol segments (which represent the cognitive activities) are embedded in writing tasks, carried out in L1 or L2 by each participant. Multilevel modelling enables us to do this hierarchy justice, while modelling the occurrence of each cognitive activity during the writing process. Previous research also indicated that polynomial models are best used in this type of analysis, due to their flexible nature (see Breetvelt, et al., 1994; Chapter 2). This enables us to model the occurrence of each cognitive activity during the writing process, because polynomials (powers of the moment) can take almost any shape, depending on the number of powers that have to be included in the model to fit the data and their values.

During the first analysis, four models were built and tested for the differences between tasks and languages, to determine which one best fit the current data. Starting point was a basic model, the *no task or language effect model (NTL)*, which only contained one fixed parameter and variances between and within individuals. This model assumes that there are no differences due to tasks or languages. Next, this model was extended to form the *task effect model (TE)*, which indicates whether there is a main effect for task, in other words, whether the mean occurrence of each activity varies between tasks. The model was built by adding 8 dummy variables to the previous model. These 8 dummies represent the 8 tasks which were included in the study. The value of each dummy is 1 if a participant completed that task and 0 if he did not.

Third, the second model was converted to form the *language effect model (LE)*, which indicates whether there is a main effect for language. In this model language was kept constant across tasks. So next to differences between tasks, differences between languages were estimated as well. We incorporated language as a fixed factor in the model, by adding an additional dummy variable; the value of the dummy is 0 if the task was completed in L1 and 1 if the task was completed in L2.

Subsequently, the LE model was transformed into the *interaction between task and language model (ITL)*, to determine whether there was an interaction effect between language and task, (i.e., whether the language effect varies between tasks). In this model the differences between languages are allowed to vary between tasks. This was done by including an additional 8 dummy variables in the model, one for each task in L2. The value of each dummy for a specific task-language combination is 1 if a participant completed that task in that language and 0 if he did not.

Comparisons of the four models indicate whether there is an effect for task, language or an interaction effect, and thus which model must be used to analyse the data.

In a second analysis we modelled the temporal distribution of the occurrence of each cognitive activity. Time was introduced as a factor in the model, because research has shown that the temporal distribution of each activity varies over the writing process. Segment number was included as an approximation of time (Van den Bergh, Rijlaarsdam, Janssen, Braaksma, Van Weijen, & Tillema, in press), due to the high correlation between total writing time and the number of protocol segments (r =. 87). So, if writers wrote for a long time, their protocols contained a lot more segments than protocols produced when they spent less time writing (see also Braaksma, et al*.,* 2004). To model the change in the occurrence of each activity over time, multivariate models were constructed for each activity in each language. As mentioned previously, we prefer to use polynomials to model the occurrence of activities over the writing process (see Breetvelt, et al*.,* 1994; Rijlaarsdam & Van den Bergh, 1996; Van den Bergh, et al., in press). In polynomials, the occurrence of each activity is modelled in terms of powers of the moment. If the number of powers differs between languages, there is a clear, significant difference due to language. But if the polynomials contain the same components in both languages, then it becomes necessary to test explicitly, by means of contrast testing, whether the temporal distributions differ significantly due to language. This is done by comparing the values of the polynomials' components (see Goldstein, 1995; Snijders & Bosker, 1999).

Finally, it was also necessary to include process-product relations in the model. This was done through the analysis of a multivariate model with process and text quality as dependent variables. The correlation between process and product was estimated at different levels: within and between individuals. Therefore, the relation between process and product was allowed to covary over time (see Van den Bergh & Rijlaarsdam, 1996; Van den Bergh et al*.,* in press).

RESULTS

This study was set up to determine whether the effect the occurrence of cognitive activities has on text quality changes as a result of writing in L2, by comparing process-product relations in L1 and L2. In this section, we will present the results of the analysis in three steps. First, we will present the model used for the analysis, followed by a comparison of the average temporal distributions of each cognitive activity (over tasks and writers) in L1 and L2. By comparing writers' own average temporal distributions to the overall average we can determine how much variation there is between writers in each language. Second, we will compare the process-

product relations in both languages and determine what the influence is of writing in L1 versus L2 on text quality. Finally, we will discuss to what extent writers' L1 approach to writing can predict how they write in L2.

Orchestration of activities in L1 and L2

To determine which model best fit the data, a pair wise comparison was carried out of the differences between the – 2 log likelihood values for each of the models. The results of this analysis are shown in Table 3. The differences between the -2 log likelihood values all have a chi-square distribution. The results clearly indicate that in each case the most complex model, the ITL *interaction between task and language model* best represented the data, for each of the four cognitive activities. This means that we must base our analyses on a model which includes differences between tasks which are allowed to vary between languages.

Results indicate that the average proportion[4] varies due to task and language and that the differences between languages are not the same for each task. In addition, the average temporal distribution over the writing process varies between the cognitive activities and varies due to language. In other words, for each activity, the differences between tasks are language dependent; the extent to which they vary between languages is different for each task. Although the differences are significant for all activities, the actual patterns of the curves in L1 and L2 might still be similar. The parameter estimates of the model, in logits, are provided in Appendix D. As these estimates are hard to interpret in logits, we converted them to proportions, which indicate the likelihood or probability that an activity will occur at specific moments during the writing process.

Table 3: Testing statistics for the differences between the four models: no task or language effect model (NTL), task effect model (TE), language effect model (LE) and interaction between task and language model (ITL)

Model	*Reading*	*Formulating*	*Planning*	*Generating*	*df*
NTL vs. TE	884.6	36.0	373.0	283.2	7
TE vs. LE	273.8	206.0	56.7	178.2	1
LE vs. ITL	715.5	373.0	2496.9	2657.5	7

The average temporal distributions per activity in both languages, averaged over writers and tasks, are presented in Figure 1. A visual comparison of the results for L1 (left) with those for L2 (right) reveals that although there are significant

[4] This is actually the logit of the average proportion. Please note: logit (P) = ln ($e^P/(1+ e^P)$)

differences (Reading the assignment: $\chi^2 = 16.96$; $df = 1$; $p < .001$; Formulating: $\chi^2 = 5.60$; $df = 1$; $p = .018$; Planning: $\chi^2 = 12.55$; $df = 1$; $p < .001$; Generating: $\chi^2 = 27.74$; $df = 1$; $p < .001$), the L1 and L2 results are in general rather similar, because the trends are the same in both languages (see Figure 1).

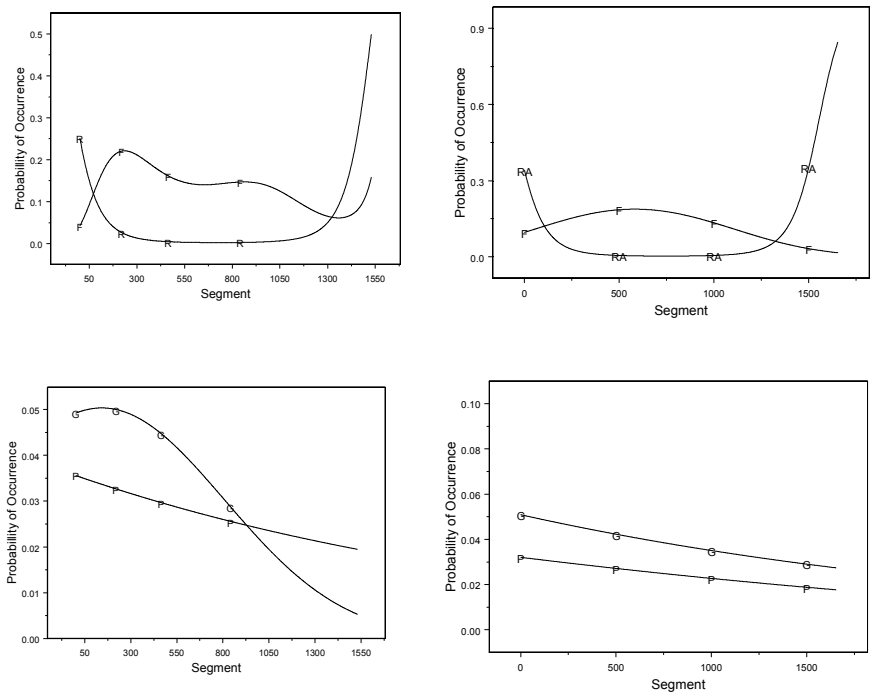

Figure 1: The mean change in the probability of occurrence of Reading the assignment (R) and Formulating (F) (top row) and Planning (P) and Generating (G) (bottom row) in L1 (left) and L2 (right).

In other words, a visual or qualitative comparison appears to indicate a similar trend in both languages (i.e., similarities), while a quantitative comparison reveals significant differences. Similarities are, for example, that Formulating has the highest overall probability of occurrence in both languages. In addition, Reading the assignment is most likely to occur at the start and end of the writing process in both L1 and L2, while Planning and Generating are rather infrequent activities in both languages. A more detailed inspection, however, reveals some differences: for example, the probability of occurrence for the two language dependent activities, Reading and Formulating, are affected by writing in L2. The probability of occurrence for Reading is somewhat larger at the start of the writing process in L2

than in L1, while the probability of occurrence for Formulating appears to vary somewhat less over time in L2 than in L1. The probability of occurrence for Generating, one of the language independent activities, also differs slightly between languages. So overall, we can conclude that, although there are quantitative differences, the average temporal distributions of these activities appear rather similar between languages.

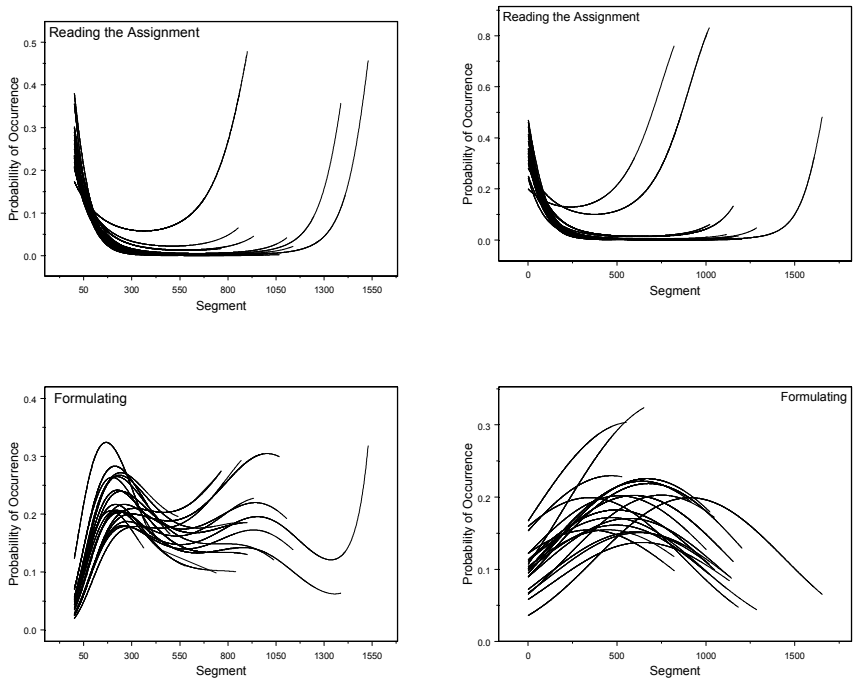

Figure 2a: The average probability of occurrence for Reading the assignment (top) and Formulating (bottom) per writer (n = 20) in L1(left) and L2 (right).

The next step in the analysis is to compare the average temporal distributions in L1 and L2 per writer, to determine how the variation in orchestration between writers varies between languages. The results for the language dependent activities, Reading and Formulating, are presented in Figure 2a. It shows that the probability of occurrence of each activity varies between writers as well as over time. For Reading the assignment, the variation between writers is largest at the start of the writing process in both languages, although it appears to be slightly larger in L2 than in L1. Differences between writers subsequently decrease and almost

disappear, with one or two exceptions. For Formulating, the variation between writers is clearly larger in L2 than L1 at the start of the writing process, although it subsequently becomes larger in L1 than in L2 near the end of the writing process.

The results for the conceptual activities, Planning and Generating, are depicted in Figure 2b. For Planning, the general picture is rather similar between languages. In both cases the variation between writers seems largest at the start of the writing process. It appears slightly larger in L2 than in L1, although this is due to only two writers, who are clearly far more likely to plan than the other writers in the study.

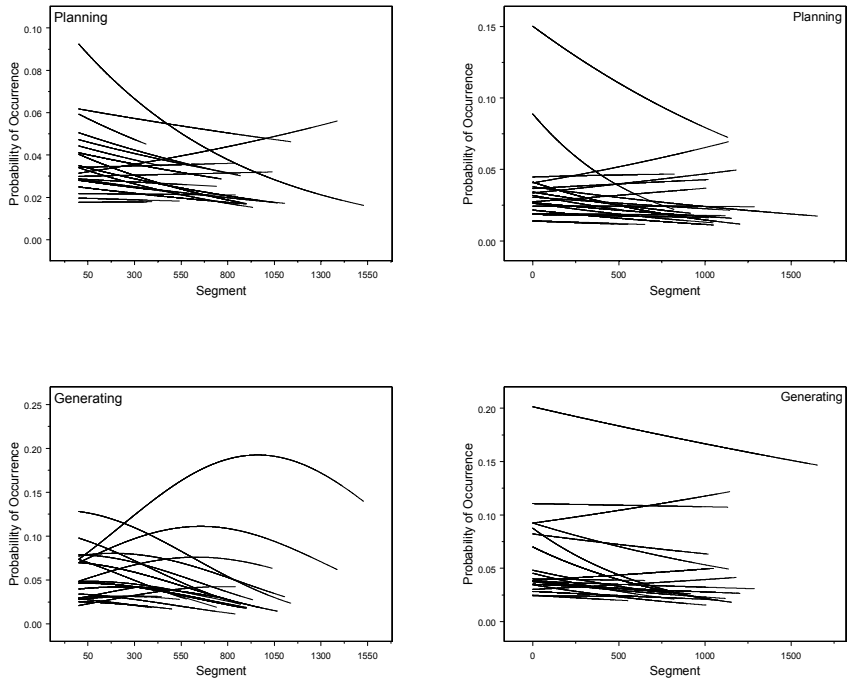

Figure 2b: The average probability of occurrence for Planning (top) and Generating (bottom) per writer (n = 20) in L1(left) and L2 (right).

Generating, finally, has a rather low probability of occurrence in L1 and L2, but the interwriter variation is clearly different in both languages. In L1, the variation is largest near the end of the writing process, due to the enormous increase in the probability of occurrence for two writers compared to the others. In L2, on the other hand, the variation between writers appears largest near the start of the

writing process, although this is again due to the behaviour of a single writer in the sample. As predicted, the orchestration of Formulating, one of the language dependent activities, appears most affected by writing in L2, although Generating appears somewhat affected as well.

Task-related variation in L1 and L2

To determine the extent to which the variation in orchestration is due to tasks or individuals, we calculated the intraclass correlations for each cognitive activity. The intraclass correlation is calculated by dividing the variation between individuals by the total variation (see Snijders & Bosker, 1999, p. 17). If this value is relatively high, then the variation in orchestration is mainly due to differences between writers. But if it is low, then the variation is mainly due to differences between tasks. Table 4 contains the intraclass correlations indicating the proportion of variation due to individuals. As the execution of each activity varies over time, we calculated the correlations for each activity at three points during the writing process: in the beginning (near segment 1), the middle (around segment 450) and the end of the writing process (near segment 900).

Table 4: The intraclass correlations indicating the proportion of between writer variation for each cognitive activity in L1 and L2

	Cognitive activities							
	Reading		*Formulating*		*Planning*		*Generating*	
	L1	L2	L1	L2	L1	L2	L1	L2
Start	.23	.75	.40	.77	.40	.68	.50	.71
Middle	.43	.62	.23	.33	.48	.61	.58	.70
End	.44	.65	.80	.73	.40	.45	.40	.58

Results indicate that in general, the proportion of the variation due to individuals is much larger in L2 than in L1. For Reading the assignment, Planning and Generating, the proportion of variance due to individuals is larger in L2 than in L1 throughout the writing process. For Formulating it is larger in L2 at the start and end of the writing process, but not in the middle. Hence, it follows that the proportion of the variation in the orchestration of activities due to task is smaller in L2 than in L1. In L1 the orchestration varies greatly between tasks, while in L2 the orchestration of each activity is much more similar across tasks. So, writers'

approach appears more uniform between tasks in L2 than in L1 for all four activities.

Process-Product Relations in L1 and L2

For the next step in the analysis, a comparison of process-product relations in L1 and L2, the correlation between the occurrence of each activity and text quality in both languages was calculated. This correlation can vary under influence of the task at hand, so we also calculated the 80% confidence intervals due to task differences. We will first compare the overall correlation with text quality in L1 and L2 and then discuss how the correlation varies between tasks in both languages. The correlations for the language dependent activities, Reading and Formulating, are presented in Figure 3a and those for the conceptual activities, Planning and Generating in Figure 3b.

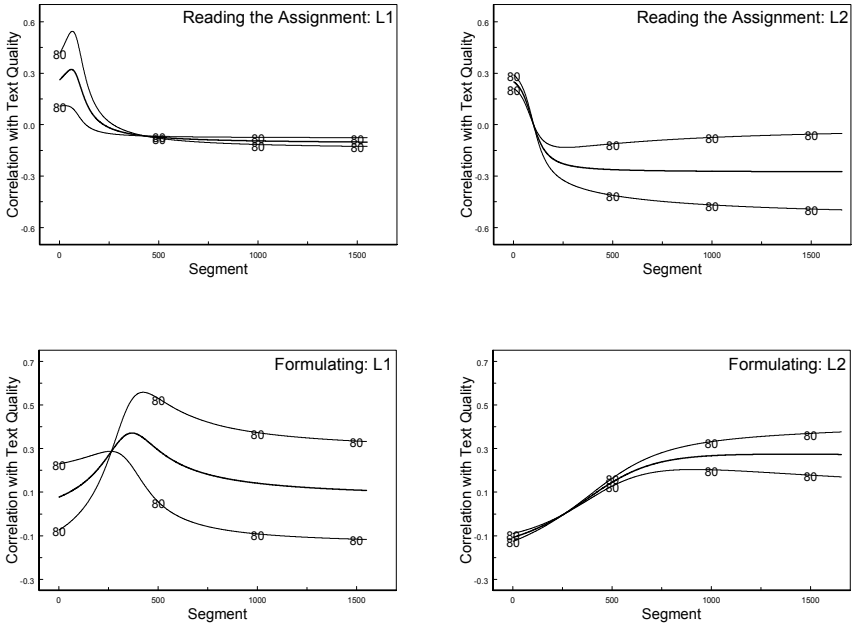

Figure 3a: The correlation between Reading (top) and Formulating (bottom) and text quality in L1 (left) and L2 (right): average (solid line) and 80% confidence intervals due to task (80).

The correlation between Reading the assignment and text quality clearly varies over time in both languages. In both cases it is highest (most positive) in the beginning, after which it decreases rapidly and becomes negative (see Figure 3a). In L1 the correlation stabilizes around -.10, while in L2 the correlation stabilizes after the

start of the writing process around -.30. Thus, in both languages writers who are likely to read the assignment at the start of the writing process are likely to write better texts than writers who are likely to read the assignment near the end of the writing process.

For Formulating, the correlation with text quality is somewhat different in both languages. Initially, the correlation is negative in both languages, after which it becomes positive. Subsequently, it peaks in L1 around .37 after which it remains positive, but decreases gradually over time to around .12 near the end of the writing process. This means that writers who focused on Formulating after the start of the writing process are likely to have written better texts than writers who focused on Formulating near the end of the writing process. In L2, however, the correlation increases over time and then stabilizes around .27 after the middle of the writing process. This indicates that in L2, writers who are more likely to concentrate on Formulating in the beginning of the writing process are likely to write poorer texts than writers who are more likely to formulate later in the writing process.

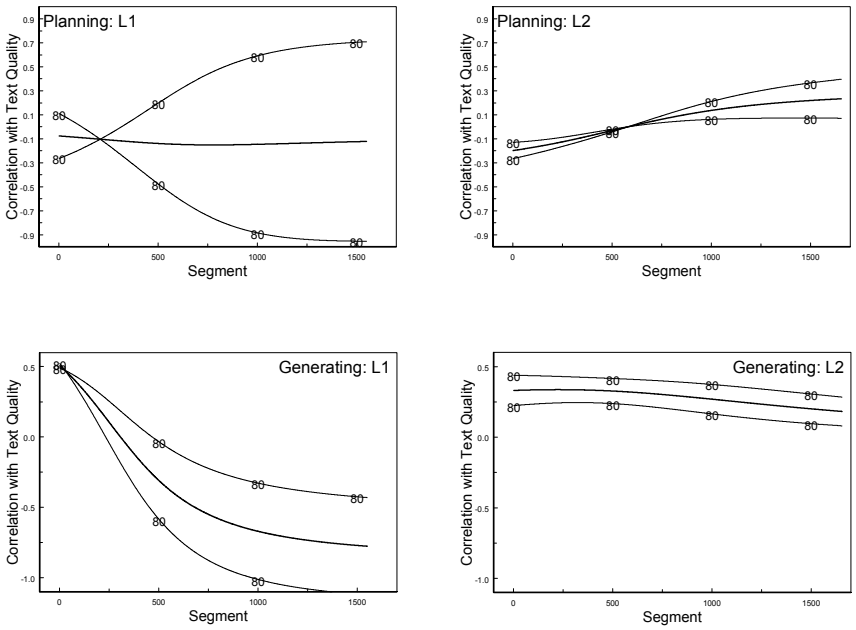

Figure 3b: The correlation between Planning (top) and Generating (bottom) and text quality in L1 (left) and L2 (right): average (solid line) and 80% confidence intervals due to task (80).

Figure 3b shows that, the correlation between Planning and text quality (solid line) clearly varies between languages and appears much stronger in L2 than in L1. In L1

the correlation is weak during the entire writing process, but in L2 it is negative at the start of the writing process, after which it becomes positive from the middle of the writing process onwards. This means that in L2, writers who are likely to plan in the beginning of the writing process are likely to write texts of poorer quality than writers who are more likely to plan during the middle or later phases of the writing process.

For Generating, the correlation with text quality also varies between languages. In L1 it is strong in the beginning (0.5), after which it becomes negative (-.75). In L2, on the other hand, the correlation is somewhat weaker in the beginning (.34), but remains positive throughout the writing process. This indicates that writers who are likely to generate at the start of the writing process in L1, are likely to have written higher quality texts than writers who generate many ideas from the middle of the writing process onwards. The same holds for Generating ideas in L2, because the correlation with text quality is also strongest in the beginning. However, Generating mainly near the end of the writing process does not seem to have the same negative influence on text quality as in L1.

Intrawriter variation in L1 and L2

As mentioned above, we determined how the correlation with text quality varied between tasks for each activity, by calculating the 80% confidence intervals for differences between tasks. If the intervals are large, then the correlation varies strongly between tasks. Results indicate that the correlation between each activity and text quality varies between tasks and over time in both languages (see lines marked with 80 in Figures 3a & 3b).

The results for Reading the assignment and Formulating are presented in Figure 3a. For Reading, the variation is largest at the start of the writing process in L1, while in L2 it is largest after the start of the writing process. This suggests that the assignment plays a far smaller role in L1 than in L2 after the start of the writing process. Figure 3a also shows that for Formulating, the correlation with text quality varies most strongly between tasks near the end of the writing process in L1 and L2.

For Planning, the correlation with text quality varies enormously between tasks in L1 after the beginning, which indicates that the correlation is extremely task dependent. In L2, it also varies between tasks, especially near the end of the writing process, but the variation appears much smaller. This could be a sign that writers must plan what they do throughout the writing process in L2 for every writing task, while in L1 whether Planning is required in later stages of the writing process depends much more on the task at hand. Finally, Figure 3b indicates that, in general, the correlation between Generating and text quality varies strongly in both languages. But an important difference between the two languages is the fact

that the correlation decreases strongly over time in L1, whereas it is more or less constant throughout the writing process in L2.

Predictive value of L1 orchestration for L2 text quality

In the final step of the analysis, we determined to what extent writers' L1 approach to writing can predict how they write in L2. To do so, we estimated the covariance between the residuals for L1 and L2 and text quality in both languages (see Figure 4). Results indicate that the extent to which writers' focus on Reading the assignment in L1 has no predictive value for text quality in L2, as the correlation varies from -.05 to 0.05 over the writing process (see Figure 4, top left).

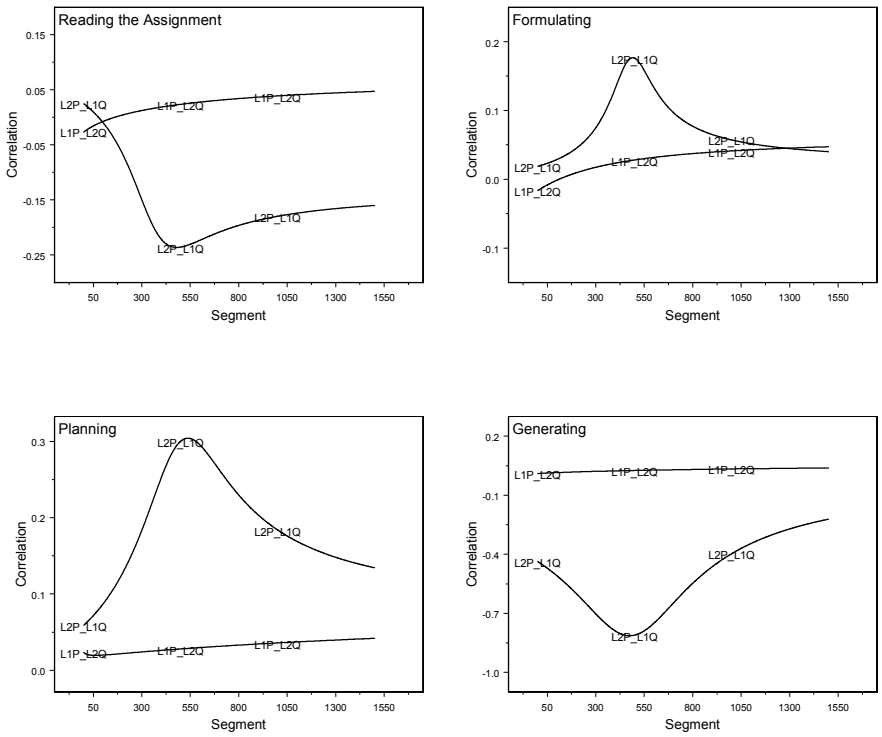

Figure 4: The correlation between L1 Process and L2 Product (L1P_L2Q) and L2 Process and L1 Product (L2P_L1Q) for each of the four cognitive activities.

Surprisingly, writers' Reading behaviour in L2 does appear to have some predictive value for L1 text quality, although only after the start of the writing process, when the correlation is moderately negative (-0.25). This suggests that if writers are likely to read the assignment frequently in English a little after the start of the writing

process, then they might have written poor quality texts in Dutch. Why? Well, the temporal distribution of Reading the assignment in L1 and L2 are positively correlated (around .65). That means that if writers are likely to read frequently at a certain point in L2 then they are also likely to do so in L1. Furthermore, if they are likely to do so at the wrong time in English (after the start of the writing process) they are also likely to do so in Dutch, which might have a negative influence on text quality in both languages.

Writers' Formulating behaviour in L1 does not appear to have a predictive value for text quality in L2. In addition, their Formulating behaviour in L2 only appears to have a very weak predictive value for text quality in L1 at best, as the correlation is only 0.18 at its peak. In general, the values are so low that it is barely possible for L1 or L2 orchestration to predict text quality in the other language. This is not surprising, considering the fact that Formulating is a very language dependent process.

The picture for Planning is similar to that for Reading; writers' Planning behaviour in L1 has no predictive value for text quality in L2 (see Figure 4, bottom left). But writers' Planning behaviour in L2 can predict L1 text quality to some extent after the start of the writing process, when the correlation is moderately positive (around 0.30). This means that if writer are likely to plan frequently after the beginning when writing in L2, then it is likely that they will also do so in L1, thus resulting in higher quality texts in L1.

Finally, the results are much stronger for Generating than for the other three activities (see Figure 4). Again, Generating in L1 does not appear able to predict text quality in L2, but Generating in L2 appears to be a strong predictor of L1 text quality. This is particularly true just after the start of the writing process when the correlation is very high (around -.80). Thus writers who are likely to generate lots of ideas in L2 at that time are very likely to have produced poor quality texts in their L1. This is due, at least in part, to the correlation between the temporal distributions for Generating in L1 and L2 (.59 to .72), as was the case for Reading the assignment. So if writers generate many ideas at the wrong time during the writing process in L2, they will probably also do so in L1, which will probably result in poorer quality texts. However, this could also be related to the language (L1 or L2) that writers generate ideas in while writing in L2, as this could also influence the quality of the text produced. If writers generate ideas mainly in L1 while writing in L2, this could be a sign of lack of L2 proficiency and thus result in a poor quality text, or it could be a coping mechanism to prevent cognitive overload and thus result in good or average quality texts.

CONCLUSION AND DISCUSSION

The aim of this study was to determine whether the effect of cognitive activities on text quality changes when writing in L2. Our first hypothesis was that the orchestration of each cognitive activity would change during writing in L2 and that the change was more likely to occur for language dependent activities than for language independent activities. Results indicate that there were significant quantitative differences in the average temporal distributions for each of the four cognitive activities in L1 and L2. Reading the assignment, for example, appears to occur somewhat more frequently in L2. In addition, Formulating and Generating appear to vary somewhat less over time in L2 than in L1. However, a qualitative comparison revealed that in general the temporal distributions of all four activities over the writing process appeared rather similar in both languages.

Second, we compared writers' average behaviour for each activity and the extent to which they varied their behaviour between tasks in both languages. For Reading the assignment and Planning, the variation was fairly similar between languages, but for Generating and Formulating, there were clear differences. In L2 the variation was largest at the start of the writing process for both activities, while in L1 it was largest near the end of the writing process. These results indicate that the temporal distributions of all four activities changed under influence of writing in L2, not just the distributions of the language dependent activities, Reading and Formulating.

Our second hypothesis was that the temporal distribution of each cognitive activity was more likely to vary between tasks during writing in L1 than in L2. The prediction was that the increased cognitive complexity of writing in L2 would result in a reduction in working memory resources available for adapting to different writing tasks, and thus result in less between-task variation in L2 than in L1. Results of a comparison of between-task variation in both languages indicated that differences were indeed generally larger in L1 than in L2, although this depends on the moment during the writing process. Conversely, the variation due to individuals was generally much larger in L2 than in L1 for all activities. So the results appear to confirm our hypothesis; there was generally more between-task variation in L1 than in L2. Or, to put it differently, writers' behaviour seems to be more uniform between tasks in L2 than in L1, particularly in the beginning of the writing process. The relative increase in between-task stability in L2 suggests that writers found writing in L2 more complex than writing in L1.

The results also appear to confirm our third hypothesis that changes in the orchestration of activities in L2 are likely to result in changes in process-product relations. Results of the analysis indicated that differences in orchestration, or more precisely in the temporal distributions of cognitive activities between languages,

appear to be related to differences in text quality, although this relation varies between activities. For example, the correlation between Reading the assignment and text quality was rather stable in L1 and L2, while the correlations between the other activities and text quality varied between languages. Planning, for example, only correlated positively with text quality near the very start of the writing process in L1, although the correlation varied enormously between tasks. In L2 on the other hand, the correlation between Planning and text quality only became positive in the middle of the writing process. This suggests that Planning played a more important role during later phases of the writing process in L2 than in L1. Perhaps writers only need to plan during later phases of the writing task during complex tasks in L1, while in L2, it seems necessary for all tasks. On the other hand, the definition of Planning in this study could provide an alternative explanation for the L2 results. Self-instructions were included as a type of Planning in this case, while in actual fact, they might play a different role during the writing process than the other subtypes of Planning: goal setting and structuring. In general, self-instructions are likely to be used most often to solve problems writers encounter during the writing process. Therefore, the increase in Planning near the end of the writing process in L2, and the subsequent positive correlation with text quality might be due to an increase in self-instructions in an attempt to get out of trouble.

It is also interesting to note that our results were remarkably similar to those of earlier research by Breetvelt, et al. (1994) on L1 writing, even though their participants were ninth graders instead of college students. For Reading the assignment and Formulating, for example, we found almost exactly the same correlation between each activity and text quality, in the beginning, middle, and end of the writing process. The results for Planning and Generating differed to some extent, which might be due, in part, to the difference in age between our groups of participants and due to differences between the tasks used in both studies. Nonetheless, Breetvelt et al.'s results still fell within the confidence intervals for task differences in our L1 data.

Finally, we investigated to what extent writers' approach to writing in L1 could predict how successful they will be when writing in L2. This is useful, because it might help explain why writers sometimes produce texts of lower quality in L2 than might be expected based on their L1 performance. Contrary to our expectations, it proved impossible to predict L2 text quality on the basis of writers' L1 orchestration. It was possible, however, to predict L1 text quality, to some extent, on the basis of writers' L2 approach, although the strength of the prediction varied strongly between activities, from very strong for Generating, to very weak for Formulating. It seems logical that the language dependent activities are not very adequate predictors of L1 or L2 text quality. However, it is surprising that only one of the more language independent activities, Generating, is a rather strong predictor of L1 text quality, while Planning is not. This is most likely due to the fact

that Planning is a very low frequent activity (see Table 2) and also because the differences in Planning behaviour are very large between tasks in L1 (see Figure 3b). Moreover, Generating in L2 is probably a far better predictor for text quality in L1, due to the fact that it occurs almost twice as often, on average, as Planning. So, we can predict L1 quality to some extent on the basis of L2 activities but not the reverse, which is probably due, in part, to the fact that writers' approach to the writing process is far more uniform or consistent in L2 than in L1.

Suggestions for further research

One of the possible limitations of this study is related to text quality assessment. Earlier research indicated that comparing text quality scores across languages was problematic (see the introduction & method sections). As a result, an attempt was made, in this study, to assess text quality in L1 and L2 in a similar way, thus making a direct comparison of scores across languages possible. Although this was not entirely successful, it did not prevent us from comparing process-product relations across languages. Thus, we were still able to determine whether or not the occurrence of a specific activity correlated with text quality in the same way in both languages. However, future research should focus on developing and testing other methods to solve this problem in a more elegant and effective manner. Possible ideas to explore include training bilingual judges to rate essays in both languages, developing a method of assessment which focuses solely on text content instead of on linguistic features, or translating all essays to the same language (L1 or L2) before assessment.

Second, as mentioned earlier, process-product relations between conceptual activities such as Generating ideas and text quality might be influenced by the language in which such activities occur during writing in L2. If writers use their L1 frequently while writing in L2, this could be a sign that they are experiencing difficulties, perhaps due to a lack of L2 proficiency. Thus L1 use during writing in L2 might have an effect on the quality of the text produced. Therefore, research is currently under way to determine what role L1 plays in L2 writing (see Chapter 6).

Finally, we cannot infer causal relations based on the results presented here. Experimental research must be carried out to test our results and hopefully prove that the relations between the occurrence of cognitive activities during writing and text quality are causal in nature. Nonetheless, the main difference between writing in L1 and L2 appears to be a decrease in variation between tasks in L2 compared to writing in L1.

APPENDIX A: Example of an assignment

Life as a student: A hard life full of stress? Or isn't it that bad?

The NSU, the National Student Union, is organising a national essay contest, especially for students. You're also taking part. You absolutely want to win. The winning essay will be printed in all the university newspapers, including the U-blad. The U-blad is read by students and employees of the university.

The subject of the essay has already been decided and was described in the U-blad as follows:
Do you ever suffer from stress? Do you sometimes feel the pressure of deadlines, compulsory attendance or lots of exams? Or do you think it is ridiculous that there are students who suffer from stress? In the media, more and more attention is being paid to subjects such as Burn-out, 'midlife crisis', RSI and other stress related symptoms. That's why the NSU wants to pay attention to this subject in a special edition of the U-blad. We want to hear from students what they think. Decide what you think and send us your response!

Assignment:
Write an essay in which you give your opinion on the question:
"Life as a student: A hard life full of stress? Or isn't it that bad?"

The essay has to meet the following requirements, set by the Jury:

1. Your essay must be (about) half a page in length.
2. You must do your best to convince your readers, readers of university magazines, of your opinion.
3. You must give arguments to support your opinion.
4. Your essay must be structured in a good and logical way.
5. Your essay must look well-cared-for (think of language use and spelling).
6. In your essay you must use at least two extracts from the 'References' (see next page). You must include these extracts in your essay in a meaningful way.

You have 30 minutes to complete this assignment.

Good luck!

References

Stress is more than just a current popular phrase; it is something that everyone, young and old can feel in different circumstances. […] Education is aimed at results and achievements these days and not at feeling good and being allowed to be yourself. Life itself is the best form of education! In order to better prepare pupils for life, it is important to pay attention to the way stress, fear and black-outs develop. With this basic knowledge, pupils will be better prepared for their future.
Source: drs B.M.G.L. Kruitz, www.heyokah.com.

"The life of others always seems nicer. Of course you are tired after a day in school. The strange thing is that many people wish that working wouldn't make them tired. You can however desire too much from life!"
Source: Ank van der Campen, "Weekblad van Leraren", October 2nd 1975.

Psychological research by the University of Amsterdam has shown that first year university students don't have an easy life. Researcher Jelte Wilcherts: "First years are more depressed than other people their age. It is also a period during which they go through large changes in their lives, which makes them more vulnerable to problems."
Source: Marloes Zevenhuizen, www.trouw.nl, April 22nd 2004.

One in ten pupils suffers from a severe type of fear of failure. They get bad grades because they are scared. Scared of failing, scared of not living up to the expectations that parents, teachers or they themselves have. They have headaches, stomach aches or heart problems. They hyperventilate or are oversensitive. The strongly competitive nature of modern society encourages fear of failure. Schools are under pressure to deliver successful students. That causes a lot of pointless stress such as fear of failure.
Source: Klasse voor Leerkrachten volume 88, October 1998.

S. Beijne, MA [a student psychologist working at the University of Leiden] estimates that about five percent of students have serious stress related problems. That's not actually that bad. "But," Beine says, "I want to emphasize that stress is a natural phenomenon. You have to learn to live with stress. A little stress is necessary if you want to do a good job."
Source: Mieneke Scheele, www.panoplia.nl, 2004.

Researchers at the VU compared students who had to meet a deadline to students who watched bloody medical documentaries. The documentary watchers had much less immunoglobin – a substance that protects people against germs – in their saliva (spit) than the deadline workers. So the researchers believe that good stress is healthy.
Source: Cicero; Leids Universitair Medisch Centrum, March 2002.

APPENDIX B: Analytical rating form

Read each essay carefully and then assign a score on each of the points below. Circle the score of your choice. The scale runs from 1 (does not meet the set criterion at all) to 5 (meets the set criterion in every way).

Part	**Score**
1. Structure	
1.1 Title The essay has a title which clearly corresponds with the content of the essay.	1 2 3 4 5
1.2 Structure The essay contains a clear division in: introduction, argumentation and conclusion.	1 2 3 4 5
1.3 Lay-out The essay is well-organized. There is a clear division in paragraphs. Paragraphs are divided by: a blank line, indenting or are started on a new line.	1 2 3 4 5
1.4 Subtopic Every paragraph has its own (sub)topic.	1 2 3 4 5
1.5 Relationships between Paragraphs There is a clear 'train of thought' between paragraphs: it is easy to determine coherence relations between paragraphs in the text.	1 2 3 4 5
1.6 Continuity Information which belongs together is presented together.	1 2 3 4 5
2. Content	
2.1 Introduction The thesis statement is presented in the introduction and (possibly) the writer's own opinion on the thesis statement is also provided.	1 2 3 4 5
2.2 Persuasion It is clear what the writer wishes to persuade the reader to believe: a choice for or against the thesis statement.	1 2 3 4 5
2.3 References The essay contains at least two quotes from the sources, which are incorporated in the essay in a meaningful way. For example, they support the argumentation or are used as an example in the introduction.	1 2 3 4 5
2.4 Refering (quoting from the sources) Quotes from the sources are correctly marked in the essay. Literal quotes (between inverted commas) and paraphrases both include references to the sources.	1 2 3 4 5

Part	**Score**
2.5 Reader Orientation The essay is easy to comprehend for readers who have not read the assignment. The writer does not, for example, refer to the assignment or to the experimental setting in the essay.	1 2 3 4 5
2.6 Reader Involvement The writer tries to engage the reader's interest by including real-life, common-place examples in the essay.	1 2 3 4 5
3. Argumentation	
3.1 Support The argumentation consists of multiple arguments, which support the writer's opinion.	1 2 3 4 5
3.2 Relevance The argumentation does not contain too much superfluous information, i.e., information which does not help support the writer's opinion.	1 2 3 4 5
3.3 Argumentation The arguments are clearly recognizable as such; for instance by the use of constructions such as *"therefore I do(not) think that ...", "I think ...", "I do (not) agree with ..."* etc.	1 2 3 4 5
3.4 Referential and Coherence relations The referential and coherence relations are clear if they are implicit, or else are marked explicitly. Examples of markings include: *therefore, so, thus, because, firstly, secondly, thirdly, subsequently,* etc.)	1 2 3 4 5
4. Conclusion	
4.1 Conclusion The essay contains a clear conclusion, which corresponds with the rest of the text and which indicates the writer's own opinion.	1 2 3 4 5
4.2 Closing It is clear that the essay is finished, for example by a brief summary of the content of the essay or by a closing statement. There are no loose ends left.	1 2 3 4 5

APPENDIX C: Structural framework for holistic rating

- For each essay, please determine to what extent it meets the requirements described in the framework and how much better or worse each essay is than the example essay on that topic. Assign each essay a score accordingly.
- The example essay is worth 100 points. If an essay is twice as good as the example essay, it is worth 200 points. If it is twice as poor, it is worth 50 points. All scores are possible. For example, an essay could be two-and-a-half times better (250 points) or worse (40 points) or three times better (300 points) or worse (30 points).

Example of Essay Structure

I <u>Introduction</u>
The thesis statement is introduced.
Optional: The writer states his opinion on the thesis statement.

A <u>Argumentation</u>
The writer is clearly for or against the thesis statement.
The arguments support the writer's opinion.

C <u>The writer's Conclusion</u>
The writer rounds off the essay with a clear conclusion.
The conclusion contains the writer's opinion (clearly for or against the thesis statement).

Clarification:

1. Compulsory order
I – A – C

2. Possible order variations
- Only arguments for or against the thesis statement;
- All arguments for the thesis statement followed by all the arguments against it;
- All arguments against the thesis statement followed by all the arguments for it;
- The arguments for and against the thesis statement are alternated;

3. Relationships
- If only arguments for or against the thesis statement are provided, then these must be linked in a clear additive sequence (e.g., *firstly, secondly, thirdly, furthermore, in addition,* etc.);
- If arguments are provided pair wise (*for, against, for, against*) then these must be presented as clearly opposing arguments (e.g., *but, pro-con, furthermore*, etc.)
- If both arguments for and against the thesis statement are provided, then the writer's own opinion (for or against) must become clear in the conclusion;
- There are clear/logical relationships present between the paragraphs;
- The relationships between the sentences within a paragraph are logical.

APPENDIX D: Parameter estimates and standard errors (s.e.) in logits for all components of the models for all four cognitive activities in L1 and L2

Fixed part of the model

Fixed part components	*Reading the assignment estimate (s.e.)*		*Formulating estimate (s.e.)*	
	L1	*L2*	*L1*	*L2*
$\beta_0 * M^0_{ijk}$	-5.195 (.327)	-5.207 (0.062)	-1.586 (0.056)	-1.487 (0.070)
$\beta_1 * M^1_{ijk}$ [5]	-0.549 (.093)	-0.501 (.000)	-0.202 (0.033)	0.047 (0.017)
$\beta_2 * M^2_{ijk}$	0.093 (.004)	0.096 (.000)	0.029 (0.005)	-0.023 (0.001)
$\beta_3 * M^3_{ijk}$			0.015 (0.001)	---
$\beta_4 * M^4_{ijk}$			-0.004 (<.001)	---
$\beta_5 * M^5_{ijk}$			0.000 (<0.0001)	---

Fixed part components	*Planning estimate (s.e.)*		*Generating estimate (s.e.)*	
	L1	*L2*	*L1*	*L2*
$\beta_0 * M^0_{ijk}$	-3.474 (0.103)	-3.571 (0.138)	-3.037 (0.136)	-3.110 (0.155)
$\beta_1 * M^1_{ijk}$ [5]	-0.040 (0.020)	-0.034 (0.021)	-0.069 (<.001)	-0.038 (0.017)
$\beta_2 * M^2_{ijk}$	---	---	-0.012 (<.001)	---

[5] This value is centred on the mean and divided by 100, which is merely a linear transformation for the sake of convenience; otherwise the estimates would be extremely small.

Random part: Interwriter variance

	Reading the assignment estimate					
	L1			*L2*		
Random part components	$S^2\ \beta_1$	$S^2\ \beta_2$	S^2*Text Quality*	$S^2\ \beta_1$	$S^2\ \beta_2$	S^2*Text Quality*
$S^2\ \beta_0$	1.531			1.694		
$S^2\ \beta_1$	0.424	0.120		0.414	0.104	
S^2*Text quality**	-0.065	-0.031	0.626	-0.289	-0.076	0.725

* = *Standardized*

	Formulating estimate					
	L1			*L2*		
Random part components	$S^2\ \beta_1$	$S^2\ \beta_2$	S^2*Text Quality*	$S^2\ \beta_1$	$S^2\ \beta_2$	S^2*Text Quality*
$S^2\ \beta_0$	0.047			0.086		
$S^2\ \beta_1$	0.015	0.017		0.002	0.005	
S^2*Text quality**	0.059	0.007	0.626	0.033	0.015	0.724

* = *Standardized*

Random part: Interwriter variance (continued)

	Planning estimate					
	L1			*L2*		
Random part components	$S^2\ \beta_1$	$S^2\ \beta_2$	S^2*Text Quality*	$S^2\ \beta_1$	$S^2\ \beta_2$	S^2*Text Quality*
$S^2\ \beta_0$	0.155			0.318		
$S^2\ \beta_1$	-0.003	0.004		0.001	0.005	
S^2*Text quality**	-0.041	-0.003	0.626	-0.019	0.019	0.723

* = *Standardized*

	Generating estimate					
	L1			*L2*		
Random part components	$S^2\ \beta_1$	$S^2\ \beta_2$	S^2*Text Quality*	$S^2\ \beta_1$	$S^2\ \beta_2$	S^2*Text Quality*
$S^2\ \beta_0$	0.370			0.430		
$S^2\ \beta_1$	0.028	0.014		0.002	0.003	
S^2*Text quality**	-0.102	-0.081	0.626	0.183	-0.003	0.725

* = *Standardized*

Random part: Intrawriter variance

	Reading the assignment estimate					
	L1			*L2*		
Random part components	$S^2\ \beta_1$	$S^2\ \beta_2$	S^2*Text Quality*	$S^2\ \beta_1$	$S^2\ \beta_2$	S^2*Text Quality*
$S^2\ \beta_0$	2.030			1.023		
$S^2\ \beta_1$	0.557	0.158		0.226	0.051	
S^2*Text quality**	-0.0001	-0.006	0.132	0.015	0.010	0.006

* = *Standardized*

	Formulating estimate					
	L1			*L2*		
Random part components	$S^2\ \beta_1$	$S^2\ \beta_2$	S^2*Text Quality*	$S^2\ \beta_1$	$S^2\ \beta_2$	S^2*Text Quality*
$S^2\ \beta_0$	0.019			0.024		
$S^2\ \beta_1$	0.002	0.003		0.002	0.002	
S^2*Text quality**	0.022	0.013	0.132	0.004	-0.005	0.006

* = *Standardized*

Random part: Intrawriter variance (continued)

	Planning estimate					
	L1			*L2*		
Random part components	$S^2\ \beta_1$	$S^2\ \beta_2$	S^2*Text Quality*	$S^2\ \beta_1$	$S^2\ \beta_2$	S^2*Text Quality*
$S^2\ \beta_0$	0.140			0.180		
$S^2\ \beta_1$	-0.009	0.006		0.013	0.007	
S^2*Text quality**	-0.048	-0.022	0.132	0.077	0.011	0.006

* = *Standardized*

	Generating estimate					
	L1			*L2*		
Random part components	$S^2\ \beta_1$	$S^2\ \beta_2$	S^2*Text Quality*	$S^2\ \beta_1$	$S^2\ \beta_2$	S^2*Text Quality*
$S^2\ \beta_0$	0.226			0.147		
$S^2\ \beta_1$	0.036	0.022		0.003	0.004	
S^2*Text quality**	0.073	0.018	0.132	0.002	0.003	0.006

* = *Standardized*

CHAPTER 5

COMPOSING EPISODES: EXPLORING THE HIERARCHICAL STRUCTURE OF L1 AND L2 WRITING

Abstract [1]

Research has shown that the writing process contains many different cognitive activities, such as Planning, Formulating and Revising. However, most studies to date have focused on individual activities in isolation, instead of on the interaction between them. One possible form of interaction is the potential influence of activities such as Planning and Evaluating on the orchestration of the writing process as a whole. Such activities appear to exert an influence over other activities, which suggests that the writing process has a hierarchical structure. It seems likely therefore, that individual cognitive activities might be nested within larger units within the writing process.
To investigate this idea, an analysis was carried out using Composing Episodes, larger units within the writing process, as the unit of analysis. The aim was first, to investigate whether the writing process can be broken down into composing episodes, and second, to determine whether activities at the start of episodes exert influence over the writing process as a whole. Results indicate that writers need fewer episodes on average when writing in L1 than in L2. In addition, Planning was the only activity which correlated positively with text quality when it occurred at episode boundaries. This supports our hypothesis that Planning appears able to guide strings of activities. In sum, the data provide evidence for the idea that writing is a hierarchically structured activity.

Keywords: Composing episodes; L1; L2; Planning; Text quality;

Since the early 1980's, the collection and analysis of think-aloud protocols has formed the basis for many writing process studies (e.g., Breetvelt, Van den Bergh, & Rijlaarsdam, 1994, 1996; Hayes & Flower, 1980; Levy & Ransdell, 1995, 1996; Raimes, 1994; Van den Bergh & Rijlaarsdam, 1999, 2001, 2007; to name a few). One of the reasons for the popularity of this method is the fact that it reveals what happens in a writer's mind during text production. Verbalizing one's thoughts during the writing process (thinking-aloud) appears to provide direct information on what happens in a writer's mind, because concurrent verbalization is in fact nothing more than a memory dump. In other words, writers verbalize what they think during the execution of the task, thereby providing insight into the cognitive activities they engage in during the writing process.

---

[1] Van Weijen, D., Van den Bergh, H., Rijlaarsdam, G., & Sanders, T. (2008). *Composing episodes: Exploring the hierarchical structure of L1 and L2 writing,* Manuscript submitted for publication.

Studies such as those mentioned above are often based on a model of the writing process first presented in the early 1980's (Flower & Hayes, 1980a; Hayes & Flower, 1980). The core components of the model are the writer's long-term memory, the task environment, and a cognitive component, which includes three main categories of cognitive activities: planning, formulating, and revising (Hayes & Flower, 1980). Subsequently, Hayes (1996) presented a revised and updated version of this model, which included a number of new components, such as the social and physical environment, and working memory, although it still contained the same cognitive activities as the previous model (e.g., planning and formulating), albeit under different headings (see Hayes, 1996).

Research has shown that writers can differ strongly in the way they employ these different cognitive activities. In addition, time, or more specifically, the moment at which activities occur during the writing process, is particularly important, because activities do not occur at random. Research on first language (L1) writing indicated, for instance, that some activities, such as goal setting (a type of Planning), are far more likely to occur at the start of the writing process than others, such as Formulating and Revising, which usually occur somewhat later (see Breetvelt, et al., 1994; Chapter 2). Although, each activity can, in theory, occur at any time during the writing process, in actual fact, there are dynamic patterns or trends in the temporal distributions of each activity (see Breetvelt, et al., 1994; Chapter 2). Furthermore, variations occur in these temporal distributions between writers and tasks, which appear to be related to variations in text quality (Breetvelt, et al., 1994; Rijlaarsdam & Van den Bergh, 2006; Van den Bergh & Rijlaarsdam, 2001; Chapter 2). So, it is not just the frequency of occurrence of activities that seems to influence text quality, but the moment at which they occur during the writing process.

Most of the research on L1 or second language (L2) writing so far has focused on the role of individual cognitive activities during the writing process, such as *planning* (e.g., Flower & Hayes, 1981a; Hayes & Gradwohl Nash, 1996; Jones & Tetroe, 1987; Manchón & Roca de Larios, 2007), *formulating* (e.g., Roca de Larios, Manchón, & Murphy, 2006; Roca de Larios, Marín, & Murphy, 2001), *revising* (e.g., Broekkamp & Van den Bergh, 1996; Stevenson, Schoonen, & De Glopper, 2006; Thorson, 2000) or on *multiple activities* (e.g., Breetvelt, et al., 1994, 1996; Van der Hoeven, 1996a; Chapter 2). But cognitive activities do not occur in isolation. The writing process is clearly a recursive process, during which writers continuously switch from activity to activity. Furthermore, there are indications that *cognitive activities influence each other* (see for example, Chenoweth & Hayes, 2001, p. 85) and, in addition, that *combinations of activities* appear to have an effect on text quality. Breetvelt, et al. (1996), for example, found that the correlation between generating and text quality *"changed dramatically"* when the fact that it often co-occurs with rereading the text produced so far was taken into account (Breetvelt, et

al., 1996, p. 19). They suggest that a single activity, such as generating or rereading the text produced so far, can fulfil different functions, depending on the moment at which it occurs during the writing process and on the activities which precede or follow it (Breetvelt, et al., 1996). This was further corroborated by subsequent research (Van den Bergh & Rijlaarsdam, 1999, 2007) which analysed how the effect of generating on text quality varies when it is directly preceded by different activities. Results indicated that pairs of activities (e.g., generating preceded by reading the assignment or by reading the text produced so far) correlate with text quality in different ways than generating does on its own. If pairs of activities co-occur within the writing process, as suggested above (Breetvelt, et al., 1996; Van den Bergh & Rijlaarsdam, 1999, 2007), then there might also be larger combinations of activities which work together. In other words, we should not analyse each cognitive activity individually and leave it at that (Breetvelt, et al., 1996, p. 19). While it has proved extremely useful to examine the effect of individual activities on text quality (see for example, Breetvelt, et al., 1994; Chapter 2), the time has come to examine the way these activities interact and influence each other (see for example, Hayes & Gradwohl Nash, 1996, p. 54; Kellogg, 2001, p. 223; Roca de Larios, Manchón, Murphy, & Marín, 2008, p. 44; Stevenson, 2005, p. 185; Rijlaarsdam & Van den Bergh, 2006; Van den Bergh & Rijlaarsdam, 2007; Van Weijen, 2006). Such a study clearly requires a larger unit of analysis, which contains multiple cognitive activities instead of a single activity.

In sum, previous research stressed the importance of taking time into account as a factor in writing process research (Rijlaarsdam & Van den Bergh, 2006; Roca de Larios, Murphy, & Marín, 2002, p. 44). The next step in writing process research, in our view, is to examine the moment at which activities occur in relation to each other during the writing process. We propose that there are larger units nested within the writing process. The question then is when do activities occur within these units? Activities which occur at the start of a unit are likely to have a different function than activities which occur later in these units. Perhaps activities which occur at the start of functional units are able to exert influence over or guide what happens in the rest of that unit of concentration.

The role of planning during writing

The first step in determining whether certain cognitive activities are able to influence larger groups of other activities within the writing process is to select likely candidates for such a role. Planning, which in this study consists of goal setting and structuring (see Table 1), is the first activity which springs to mind. Previous research has often stressed the important role Planning plays within the writing process and its possible influence on text quality (Flower & Hayes, 1980a, 1981a; Hayes & Flower, 1980; Van den Bergh & Rijlaarsdam, 2001; Van der Hoeven, 1999). Flower and Hayes (1980a) distinguish between three main types of

planning: a main category *Plans to Do* and two subcategories: *Plans to Say* and *Plans to Compose*. Plans to Do are general, rhetorical plans for dealing with the rhetorical problem: *"I have to write an essay on this topic"*. Plans to Say mainly deal with the content of the text: *"What did I want to write?"* or *"This paragraph is finished!"*. Plans to Compose on the other hand focus on the execution of the writing process: *"I think I'll reread the sources again."* or *"That's not exactly the word I was looking for, but I'll use it anyway."* They suggest that these three types of plans are *"hierarchically related to one another"* (Flower & Hayes, 1980a, p. 44). In a subsequent study, they also distinguish between three kinds of planning, but with different names: *Plans for generating ideas*, which resemble plans to say, *Plans for producing a paper,* which seem related to plans to do and *plans controlling the composing act*, which resemble plans to compose (Flower & Hayes, 1981a, p. 42). Thus there are clearly different types of planning, which seem to perform different functions during the writing process (Flower & Hayes, 1980a; 1981a).

By contrast, research has also shown, that on-line planning (i.e., planning during the writing process) is a rather infrequent activity (e.g., Ellis & Yuan, 2004, 2005; Penningroth & Rosenberg, 1995; Raimes, 1994; Van den Bergh & Rijlaarsdam, 2001; Chapter 2). Furthermore, *"analysis failed to reveal any advantage for writers who devoted a larger proportion of their time to planning over those who devoted a smaller proportion"* (Hayes & Gradwohl Nash, 1996, p. 54). In other words, the amount of attention devoted to planning does not appear to matter. Indeed, research has confirmed that it is not the amount of planning that matters, but the moment at which it occurs during the writing process (see Breetvelt, et al., 1994; Chapter 2). The intriguing question, then, is how Planning can have a positive effect on text quality, despite its infrequent nature. Perhaps it influences text quality indirectly, because it influences the way in which the writing process unfolds. It seems likely, that Planning guides or orchestrates the execution of other cognitive activities and thus the writing process as a whole (see for example, Flower & Hayes, 1981b, Hayes & Gradwohl Nash, 1996; Manchón & Roca de Larios, 2007; Sasaki, 2002; Van Weijen, 2006; Zimmermann, 2000). In other words, the presence of Planning activities might indicate that writers are able to monitor their actions during the writing process (see also Hayes & Flower, 1980; Oostdam & Rijlaarsdam, 1995).

Results of a relatively small-scale qualitative study (Van Weijen, 2006) indicated that analysing protocols in terms of *composing episodes* (Flower & Hayes, 1981b) could help identify ways in which cognitive activities influence each other during the writing process. These episodes appear to be units of concentration, which occur during the writing process. They each contain multiple occurrences of cognitive activities and are often initiated by Planning comments (see Flower & Hayes, 1981b). In other words, composing episodes seem useful, because they contain other activities and because Planning appears to play an important role within them. Therefore, composing episodes were chosen as the unit of analysis in

the present study. Studying the occurrence of composing episodes in the writing process might tell us something about the way writers break the complex writing task down into smaller, more manageable units (see Ellis, 2005; Manchón & Roca de Larios, 2007). Second, the use of composing episodes (Flower & Hayes, 1981b) might provide evidence for the effect Planning has on other activities:

> *"The composing process has an episodic pattern of its own which is not dictated by the patterns of the text. Writers appear to work in composing "episodes" or units of concentration which are organized around a goal or plan. Understanding the overall architecture of these episodes and the logic which begins and ends them will, we think, tell us a great deal about how writers combine planning and text production"* (Flower & Hayes, 1981b, p. 242).

If Planning really guides the execution of the writing process, then we expect the occurrence of Planning at the start of episodes to correlate with text quality. In other words, if writers have a large proportion of episodes which are initiated by Planning, then the resulting texts are likely to be of greater quality than if they have proportionally fewer episodes which begin with Planning. Finally, we wish to investigate possible differences in Planning behaviour in L1 and L2. Less research has been done to date on the role of Planning in L2 than in L1 (see Ellis & Yuan, 2004), although earlier research did indicate that writers generally plan less in L2 than in L1 (Jones & Tetroe, 1987; Silva, 1993). In addition, writing in L2 is a complex activity with a larger cognitive load in general than writing in L1. Working memory capacity is limited and writing is a cognitively complex process (Kellogg, 1994, p. 17; 2001), regardless of the language in which it is carried out. Working in L2 increases task complexity even further, and is therefore likely to result in a decrease in attention available for conceptual processes such as Planning. Research by Jones and Tetroe (1987) suggests, however, that even though writers appear to plan less in their L2, the effectiveness of their Planning behaviour is comparable in both L1 and L2. Hence, we predict that Planning might occur less frequently in L2 than in L1, but that if it occurs at episode boundaries it will correlate positively with text quality, as in L1.

Research questions

This study is guided by two main research questions: *(1) Can the writing process be broken down into composing episodes? And, if so, which cognitive activities occur at the start of episodes?* and *(2) Do activities at the start of composing episodes appear to exert influence over the writing process?* In other words, are activities at the start of episodes related to text quality?

In this study, we have several hypotheses regarding the effect of language. Our first hypothesis is that the writing process can be broken down into composing

episodes in both languages. However, we expect that composing episodes will generally be longer in L1 than L2, because we think writers' concentration curves will be longer, on average, in L1. In addition, if episodes are longer in L1, then it is likely that there will be fewer episodes, on average, in each protocol in L1 than in L2. We also expect the number of episodes and their length (i.e., the number of segments they contain) to vary between writers in both languages. Furthermore, we predict that Planning will be one of the most frequent activities to occur at episode boundaries.

With regard to the second question, we predict that activities which occur more often, proportionally, at episode boundaries than within episodes are able to influence the orchestration of the writing process in both languages. This influence will be indicated by the fact that activities which occur frequently at episode boundaries will correlate positively with text quality in both languages. We predict that Planning will be the most influential activity, but it is possible that other conceptual activities, such as evaluating also play a similar role. Finally, in line with earlier research (Jones & Tetroe, 1987), we expect Planning to occur more often during the writing process in L1 than in L2, although we expect the influence of Planning on the writing process as a whole to be similar in both languages.

METHOD

Design

Nineteen participants[2], first year BA English students, wrote two short argumentative essays each, one in their L1 (Dutch) and one in their L2 (English). Participants were told that they had to write their essays for an essay contest organised by the local University Newspaper. They were told that the goal was to win the essay contest and they were provided with background information on the topic of the assignment, in the form of six short sources, of which two or more had to be integrated in the text. They were asked to provide their opinion and convince their readers of their point of view. The two topics they had to write about were: *"Education in English: good idea or bad suggestion?"* and *"Everybody should automatically be a donor, unless they explicitly register against it: good or bad idea?"* Half of the participants wrote about the first topic in Dutch and the second topic in English, while the other half of the group did the reverse. They wrote their texts in L1 and L2 in random order. An example of an assignment and its sources can be found in Appendix A. Participants wrote their texts individually during two separate sessions with the researcher. They had to write each text within 30 minutes and were trained to work under think-aloud conditions before the first assignment was handed out.

[2] Initially, there were twenty participants, but one participant was unable to complete her L1 task due to technical difficulties, and was thus excluded from the analysis.

During each writing session, audio- and video-recordings were made while they wrote.

Data analysis

During the first phase of the analysis, the audio- and video- recordings of each writing session were transcribed and segmented to form think-aloud protocols. The coding scheme used included eleven categories, one for each cognitive activity (see Table 1). It was originally based on Hayes and Flower (1980) and used extensively in previous research (see for example, Breetvelt, et al., 1994; Couzijn, Rijlaarsdam & Van den Bergh, 2002; Van den Bergh & Rijlaarsdam, 1999; Van der Hoeven, 1996a, 1999; Van Weijen, 2006; Van Weijen, Van den Bergh, Rijlaarsdam & Sanders, 2005a, 2005b; Chapter 2).

Table 1: Overview of the eleven main categories included in the coding scheme

Cognitive activity	*Description and protocol example*
(Re)-Reading the assignment	Reading the assignment and/or the sources
Planning: Self-instructions	Instructions the participant gives himself regarding the next step in the writing process (e.g., *"I need to decide how to start"*)
Planning: Goal setting	Paraphrasing task demands in the assignment or setting new goals (e.g., *"This text is going to be about..."*)
Planning: Structuring	Selecting and ordering ideas or making a rough outline (e.g., *"First the pro's, then the con's and then my own opinion!"*)
Generating ideas	Paraphrasing ideas from the assignment or generating new ones: (e.g., *"Us humans we're very curious and we like to gossip!"*)
Metacomments	Reflections on the writing process as a whole or comments on the assignment and sources (e.g., *"I had the perfect sentence in my head"*)
Pausing	Silence or interjections (e.g., *um, ehhh*)
Formulating	Text production or typing the actual words on the computer.
Rereading own text	Rereading short or longer segments of the text produced so far
Evaluating own text	Evaluating the text produced so far at word, sentence or text level (e.g., *"No I'm not gonna say this"*)
Revising own text	Revising the text produced so far at word, sentence or text level (e.g., erases 'them' and types 'people')

Each protocol was first divided into segments, each containing a single cognitive activity, by two separate groups of six coders, one for L1 and one for L2. The interrater reliability between coders was relatively high in both languages (Cohen's kappa: L1 = .84, L2 = .95). Subsequently, in a second coding phase, protocol segments were combined to form larger units: *Composing episodes* (see next section).

We chose to focus our analysis of activities which occur at the start of episodes on those cognitive activities which might influence what happens within episodes: Planning, Generating, and Evaluating. For the purpose of this analysis Planning was split into two categories: Self-instructions (Breetvelt, et al., 1994; Rijlaarsdam, 1986) and a Planning category containing goal setting and structuring. We decided to define Planning as the occurrence of goal setting and structuring activities, because they are both activities which have the potential to guide or influence the writing process in some way (see also Hayes & Gradwohl Nash, 1996, p. 33). In addition, we decided to include self-instructions as a separate category, for two reasons. First of all, self-instructions are less likely to direct the writing process in the same way that other Planning activities might do, because they can occur for conflicting reasons. On the one hand, the occurrence of self-instructions can be a sign of metacognitive control over the writing process, but on the other hand, they can also occur when writers are experiencing trouble while writing. Second, earlier research indicated that self-instructions occur far more frequently than goal setting and structuring (see for example Chapter 3). This means that if these three activities are analysed in combination, the influence of goal setting and structuring would probably be overshadowed by the frequent occurrence of self-instructions, and would thus be very hard to determine.

Contrary to other studies, we decided to treat Generating as a separate activity, rather than a sub-process of Planning, because it does not appear to interact with the other activities in the writing process in the same way Planning does; it focuses primarily on the content of the writing process, not its orchestration. This is in line with Hayes' (1996) revised model of the writing process, which, in our view, separates Planning from Generating by including the first in the reflection component and the second in the text production component. Finally, we decided to examine the occurrence of Reading the assignment and Reading own text (i.e., the text produced so far) because writers who run out of ideas might fall back on reading as a strategy for generating new ideas. These two activities might therefore occur frequently at episode boundaries if participants were, for some reason, unable to plan.

Composing episodes

In writing process research the segmentation criteria for the analysis of think-aloud protocols are often not provided (see Roca de Larios, et al., 2002). The unit of analysis chosen to analyse the data, therefore, also remains somewhat unclear. The

mainstream idea appears to be that the unit of analysis is a single occurrence of a single cognitive activity (see for example, Braaksma, Rijlaarsdam, Van den Bergh, & Van Hout-Wolters, 2004; Breetvelt, et al., 1994, 1996; Hayes & Flower, 1980; Van den Bergh & Rijlaarsdam, 1999, 2001; Van der Hoeven, 1996a; Chapter 2). But others defined their unit of analysis differently, for example as: a section of text produced so far (Chenoweth & Hayes, 2001), one *"focus of attention"* (Penningroth & Rosenberg, 1995, p. 196), a single second of the writing process containing a single activity (Levy & Ransdell, 1995, 1996) or they segmented their data on the basis of pauses that occur during the writing process (e.g., Cumming, 1989; Schilperoord, 1996; Uzawa, 1996).

In previous research, we defined the unit of analysis as a single occurrence of a single activity, as described above (see Chapter 2). But to study the influence of a single activity over a large number of other activities, a larger unit of analysis is obviously required in the present study. To this end, Breuker, Elshout, Van Someren, and Wielenga (1986) proposed a dual coding phase: an initial analysis of single activities, followed by an analysis of larger elements in a later phase, or vice-versa. The major benefit of such a two-step coding phase is that it provides insight into the larger units present in think-aloud protocols, while still making an analysis of the activities within those units, at the segment level, possible. Therefore, we suggest first coding protocol segments in the traditional manner (containing single activities), followed by a second coding session during which these segments are combined into larger units, composing episodes. An analysis using composing episodes could provide information on the influential role activities such as Planning play during writing.

Composing episodes were defined as: *"relatively autonomous units within the writing process"* (Van Weijen, 2006, p. 154). The division of the think-aloud protocols into composing episodes occurred on the basis of indications in the protocols themselves. There are several indications that a current episode is finished and a new one is about to begin. First of all, a boundary can occur when a writer clearly indicates that a larger portion of the text such as a paragraph has been completed: *"Next paragraph!"* or *"This was my introduction."*. A second indication that a new episode is about to begin, is an abrupt transition from one process to the next, such as a transition from formulating or revising to rereading the assignment. Such a transition signals that writers have shifted their attention from text production to information processing (see also Flower & Hayes, 1981b, p. 238). A third and final indication of an episode boundary is a long pause between protocol segments. In this study, a long pause was defined as a pause of more than six seconds, or a combination of several short pauses and/or hesitations (*ehm, uh*) in rapid succession, which add up to seven seconds or more. Such a pause does not provide information about what a writer is thinking at that particular moment, but does indicate that the flow of the writing process is disrupted for some reason.

Therefore, we see it as an indication that writers' attention has been diverted and that they have momentarily lost their focus or concentration. Moreover, pauses are often followed by comments or actions which indicate that a writer is starting a new *"unit of concentration"* (Flower & Hayes, 1981b, p. 234), which is in fact a new composing episode.

The 38 think-aloud protocols were divided into composing episodes by a single researcher (the first author). To determine the reliability of this coding procedure, the procedure was repeated by the same coder after a four week delay. Initially, the agreement between the two coding sessions was moderate (Cohen's Kappa: .50). However, when the definition of agreement was extended to include cases in which the boundaries fell in adjoining segments (i.e., their boundaries were a single segment apart), the level of agreement between the two sessions increased enormously (Cohen's Kappa: .96). This seemingly large difference in the coding of boundaries in the two sessions is not as problematic as it may seem. First of all, Flower and Hayes (1981b) reported that not all boundaries between composing episodes were very clear: *"Although many boundaries are sharply defined points, some are one to three clause transitional units containing false closures, false starts and metacomments"* (Flower & Hayes, 1981b, p. 237). The same applied to most of the cases in which the boundaries differed between the first and second coding sessions in this analysis. Second, Flower and Hayes found that segments near boundaries often contained signal words or phrases such as *"ok"*, and *"let's see"* (Flower & Hayes, 1981b, p. 237). The same holds, again, for our data. Writers frequently used signal words near boundaries, which made the choice of the exact segment for the boundary somewhat difficult. For example, *"let's see"* can be interpreted as a self-instruction, if it is preceded by a text production phase or a pause and is directly followed by reading the assignment or the text produced so far. In such a situation it could be considered the start of a new composing episode. However, if it occurs before a short pause which is followed by a Planning comment such as *"What shall I write next?"* then the boundary is less clear and depends on the importance one places on a comment such as *"let's see"* in comparison to a more elaborate Planning comment, made in close proximity. So, although the coder appeared to vary somewhat in the way she determined the exact location of the boundaries between episodes, in general she was consistent in defining the number of episodes per protocol and their average length. Finally, the coder determined the definite boundaries, by examining all the occurrences which differed between the first and second coding sessions so that the rest of the analysis could be carried out.

Text quality

The quality of the texts was assessed in L1 (N = 19) and L2 (N = 19) by two groups of five raters, one for L1 and one for L2. These two groups each assessed the texts twice: once analytically and once holistically. In both languages the analytical rating

of the 19 texts was performed first, using an analytical rating form, with four main categories: *Structure, Content, Argumentation* and *Conclusion*. The holistic assessment in L1 and L2 was carried out two weeks later. Each rater was provided with a benchmark essay (worth 100 points) for each topic and had to decide how much higher or lower in quality each essay in the sample was than the benchmark essay; scores were then assigned accordingly. For both methods, the reliability was acceptable in L1 (analytical: $\alpha = .78$, holistic: $\alpha = .78$) and L2 (analytical: $\alpha = .68$ holistic: $\alpha = .88$). In addition, the correlation between the two ratings was relatively high both in L1 ($r = .68$, or .87 when corrected for attenuation) and L2 ($r = .69$; corrected for attenuation, $r = .89$). Thus a single score for text quality was calculated for each text in L1 and L2, based on the combined scores from the two different assessments.

RESULTS

The main aim of this study was to examine whether the writing process can be broken down into composing episodes, and also to determine which cognitive activities occur at episode boundaries in L1 and L2. Furthermore, to answer the second research question, we wanted to determine whether activities at the start of composing episodes are related to text quality and thus whether they appear to exert influence over the writing process as a whole.

Table 2: Average number of episodes per think-aloud protocol in each language

		Range	
Language	*Average number of episodes per protocol (sd.)*	*Min.*	*Max.*
L1 Dutch	24.63 (10.70)	10	47
L2 English	30.53 (11.06)	10	56
Overall	27.58 (11.14)	10	56

Composing episodes in L1 and L2

In the first step of the analysis, the protocols were divided into episodes (see Table 2). Results show that on average L1 protocols contained nearly 25 episodes, while L2 protocols contained almost 31 episodes, a significant difference ($\chi^2 = 4.59$; $df = 1$; $p = .03$). In other words, writers needed fewer episodes, on average, in L1 to complete a text than in L2. Contrary to our expectations, the length of these episodes, calculated as the number of segments each episode contains, did not

differ significantly between languages (L1: 47.03 (3.8); L2: 43.50 (2.3)). So the number of cognitive activities that a single episode contains does not vary between languages. However, we did find a significant difference in the variation in episode length between languages; episode length varies more strongly between writers in L1 than in L2 (L1: S^2 = 274.6 (*se* = 89.6); L2: S^2 = 99.3 (*se* = 32.5); χ^2 = 3.79; *df* = 1; *p* = .026 (1-sided)). Writers appear to alternate between longer and shorter episodes far more in L1 than in L2. This suggests that writing in L1 might have a more hierarchical structure, while L2 writing might be more sequential or continuous in nature. In L1, longer episodes, during which multiple sentences are written, might be interspersed with shorter ones, if writers get sidetracked or need to solve smaller, subordinate problems. For example, writers might make a plan in L1, attempt to execute that plan, but encounter a problem. It is likely that they will then attempt to solve that problem first, before returning to the execution of the plan they had initially. Finally, episode length is generally somewhat more stable in L2 than in L1. This is in line with earlier research (Chapter 3) which showed that, in general, writers' behaviour was more uniform across tasks in L2.

Table 3: Examples of activities which occur at episode boundaries

Cognitive activity	*Protocol segment (Participant, task)*
Planning: Self-instructions	*"well then I should start writing something I guess, that sounds like a good idea"* (10, Education)
Planning	*"let me see whether there's anything else I can add to this"* (9 Education)
Generating ideas	*"yes I think that I will argue for compulsory automatic donor registration after all"* (9, Organ donation)
Evaluating	*"but that is not it really my point, because it was about bilingual Havo VWO"* (.) "*so then I'll just delete this bit"* (3, Education)

(.) = 1 second pause

Cognitive activities at episode boundaries

In the second step of the analysis, we determined how often each activity occurs at episode boundaries. For this purpose, we defined an episode boundary as the first segment of an episode. This means that we calculated how often each activity occurs as the first segment in an episode. Because there are far fewer episode

boundaries than segments within episodes in each protocol, we calculated how often an activity occurs at episode boundaries as a proportion of the total number of episode boundaries in each language (L1: *N* = 471; L2: *N* = 579). In addition, we calculated how often each activity occurred within episodes as a proportion of the total number of segments within episodes in each language (L1: *N* = 13915 segments; L2: *N* = 15869 segments). As mentioned in the method section, the analysis focused on the following activities: Planning, Generating, Evaluating, Reading the assignment and Reading own text. Remember, for the purpose of this analysis, Planning was split into two categories: Self-instructions and a separate Planning category, containing goal setting and structuring. Examples of each activity at the start of episodes (except Reading the assignment or own text) are presented in Table 3. These examples show that each of the activities we studied occurs at episode boundaries, although they all occur with different frequencies. Boundaries that were not marked by these activities were often marked by pauses.

Table 4: The proportion of occurrence of each cognitive activity at episode boundaries and within episodes in L1 and L2

Cognitive activity	*Language*	*Mean at boundaries (sd)*	*Mean within episodes (sd)*	*Position effect*	*Language effect*	*Interaction effect*
				(*df 1* = 1; *df 2* = 18)		
Reading the assignment	L1	.084 (.038)	.031 (.014)	F=124.45	F=10.35	F = 5.68
	L2	.134 (.055)	.040 (.015)			
Reading own text	L1	.109 (.101)	.019 (.013)	F = 21.65	n.s.	n.s.
	L2	.095 (.097)	.013 (.009)			
Planning	L1	.049 (.047)	.005 (.006)	F = 26.76	F = 5.21	n.s.
	L2	.022 (.030)	.002 (.007)			
Planning: Self-instructions	L1	.229 (.146)	.016 (.011)	F = 66.52	n.s.	n.s.
	L2	.223 (.137)	.016 (.012)			
Generating	L1	.082 (.077)	.043 (.034)	F = 33.86	n.s.	F = 6.25
	L2	.128 (.088)	.037 (.024)			
Evaluating	L1	.042 (.042)	.023 (.014)	F = 6.71	n.s.	n.s.
	L2	.040 (.044)	.021 (.012)			

n.s. = non-significant

The results of the subsequent analysis, presented in Table 4, indicate that, for all the activities included in the analysis, the proportion of occurrence is significantly higher at episode boundaries than within episodes in both languages. We found large main effects for Reading own text (F (1, 18) = 21.65; $p < .001$; $\eta^2 = 0.55$), Planning (F (1, 18) = 26.76; $p < .001$; $\eta^2 = 0.60$) and Self-instructions (F (1, 18) = 66.52; $p < .001$; $\eta^2 = 0.79$) and a medium effect for Evaluating (F (1, 18) = 6.71; $p < .05$; $\eta^2 = 0.27$) In addition, Planning was the only activity that occurs significantly more often in L1, on average, than in L2, both at episode boundaries and within episodes (F (1, 18) = 5.21; $p = .035$; $\eta^2 = 0.23$). Finally, we found medium sized interaction effects, for Reading the assignment (F (1, 18) = 5.68; $p = .028$; $\eta^2 = 0.24$) and Generating (F (1, 18) = 6.25; $p = .022$; $\eta^2 = 0.26$). For both activities, the difference between L1 and L2 is larger at episode boundaries than within episodes.

Table 5: Examples of planning at episode boundaries and within episodes

Cognitive activity	*At episode boundary*	*Within episodes*
Planning: Goal setting	*"I have to use at least one other quote"* [from the sources]	*"A title!"*
	"I'm trying to decide what my main point will be"	*"I just realised that I have to start writing or I'll run out of time".*
Planning: Structuring	*"Ok. Now you must back this up with arguments."*	*"I'll use that as a subtitle."*
	"I think I'll just quote that whole piece, that whole reference and start my text that way"	*"Well, I think I've extracted all the sources which are in favour"*

But do any of these activities influence what happens within composing episodes? Table 4 indicates that Planning is a very infrequent activity. Yet it occurred more often proportionally at episode boundaries in both languages than within episodes. And, as expected, it also occurred significantly more often in L1 than in L2. Some examples of Planning activities at episode boundaries and within episodes are provided in Table 5, which indicate that Planning at episode boundaries and Planning within episodes are qualitatively different things. Planning at episode boundaries appears to indicate the presence of Plans to do and Plans to compose, which are both related to control over the writing process, while Planning within

episodes appears to indicate the presence of Plans to say, related to generating content and short term goals. But all the other activities in the analysis also occurred significantly more frequently at episode boundaries than within episodes. This suggests that they might all play an important role, influencing or guiding the writing process, as we predicted. If this is true, then we expect the occurrence of each activity at episode boundaries to correlate positively with text quality.

Composing episodes and text quality

For the next step in the analysis, we divided our participants into two groups for each language: good writers, with above average text quality scores, and weaker writers, with below average text quality scores. Participants were assigned to these groups based on their average text quality scores in each language, so there were four groups in total: Group 1 consisted of three writers who were good writers in both languages, Group 2 consisted of three writers who were good writers in L1 but relatively weak writers in L2, Group 3 contained seven writers who were weak in L1 but good writers in L2, and Group 4 contained six writers who were weak writers in both languages. This shows that nine writers were relatively stable, as they were relatively good or weak writers in both L1 and L2. The other ten writers, on the other hand, were good writers in one language and weak writers in the other. Subsequently, for each language, we determined whether these groups differed in the extent to which they engaged in each activity at episode boundaries or within episodes.

Results for L1 indicate that, in general, good and weaker writers did not differ in this respect (see Table 6). There were no significant differences between good and weaker writers for Reading the assignment, Reading own text, Generating and Evaluating. Although we thought Generating and Evaluating could also have a positive influence on what happens within composing episodes and thus perhaps influence text quality in a positive way, this was clearly not the case. Good writers and weaker writers did not differ significantly in their Generating (F (1, 18) = .20; p = .66; η^2 = .012) or Evaluating behaviour (F (1, 18) = .60; p = .45; η^2 = .034).

Writers only differed significantly in the way they applied two activities: Planning and Self-instructions. On average, both good and weaker writers plan more often at episode boundaries than within episodes (F (1, 18) = 27.32; p < .001; η^2 = .62). In addition, we found a medium sized interaction effect, which indicates that good writers plan relatively more at episode boundaries than weaker writers (F (1, 18) = 5.19; p <.001; η^2= .23). In other words, Planning at episode boundaries is positively related to text quality in L1. For Self-instructions, we also found a large effect; on average, both good and weaker writers used significantly more Self-instructions at episode boundaries than within episodes (F (1, 18) = 38.34; p < .001; η^2 = .70). In addition, weaker writers use significantly more Self-instructions at episode boundaries than good writers (F (1, 18) = 8.37; p = .01; η^2 = .33). This

indicates that the occurrence of self-instructions at episode boundaries is negatively correlated to text quality in L1.

Table 6: The proportion of occurrence of each cognitive activity at episode boundaries and within episodes in L1 and L2 for good and weak writers

		Weak writers (below average)		*Good writers (above average)*	
Cognitive activity	*Language*	*Within episodes*	*At episode boundary*	*Within episodes*	*At episode boundary*
Reading the assignment	L1	.029	.075	.033	.105[a]
	L2	.044	.132	.036	.136[a]
Reading own text	L1	.019	.100	.018	.130[a]
	L2	.012	.090	.014	.100[a]
Planning	L1	.006	.036	.002	.078[a, b]
	L2	.004	.010	.000	.032[a, b]
Planning: Self-instructions	L1	.017	.283	.015	.112[a, b]
	L2	.017	.266	.016	.185[a, b]
Generating	L1	.045	.087	.038	.070
	L2	.036	.137	.038	.121
Evaluating	L1	.026	.045	.017	.035
	L2	.024	.048	.017	.032

a: main effect of within episodes vs. at episode boundaries (p <.05);
b: interaction with text quality (p <.05)

The results for L2 were similar. In general, good and weak writers did not differ significantly in the extent to which they engaged in each activity, whether at episode boundaries or within episodes (see Table 6). Again, good and weaker writers did not differ significantly in their Generating (F (1, 18) = .07; p = .80; η^2 = .004) or Evaluating behaviour (F (1, 18) = .32; p = .32; η^2 = .06). As in L1, writers only differed in the way they planned. Both good and weaker writers plan more often at episode boundaries than within episodes (F (1, 18) = 9.98; p = .006; η^2 = .37), but good writers plan significantly more at episode boundaries than weaker writers (F (1, 18) = 4.95; p = .04; η^2= .23). So Planning at episode boundaries also appears to be positively related to text quality in L2, although the effect size for the difference

between Planning at episode boundaries and Planning within episodes appears smaller than in L1. Finally, the use of Self-instructions was not significantly related to text quality in L2. Although weaker writers appear to use more Self-instructions at episode boundaries than good writers, this difference was not significant (F (1, 18) = 1.89; p = .188; η^2 = .10). Thus the negative influence of Self-instructions at episode boundaries found in L1 no longer holds when writing in L2.

CONCLUSION AND DISCUSSION

This study was set up to determine whether certain cognitive activities, such as Planning, influence other cognitive activities during the writing process. Therefore, we analyzed our data in terms of composing episodes. We can draw a number of conclusions, on the basis of our findings. First of all, results indicated that writers need significantly fewer composing episodes to produce a text in L1 than in L2. This lends some support to our hypothesis that writers' concentration curves would be longer in L1 than in L2, even though the difference in episode length was not significant between languages. However, episodes did vary in length to a greater extent between individuals in L1 than in L2. This indicates that writers alternate more between longer and shorter episodes in L1 than in L2. This could be due to the increase in cognitive load that writing in L2 entails, as earlier research also indicated that writers vary their behaviour less between tasks when writing in L2 than they do in L1 (see Chapter 3). Writing in L2 appears to increase writers' cognitive load, which makes it impossible for writers to adapt to changing task demands to the same extent as in L1 writing. In other words, it seems likely that writers have more working memory resources available in L1, which enables them to deviate from their main goal or plan while writing if a problem arises which demands immediate attention. The question remains, however, whether episodes are more strongly hierarchically related in L1 than in L2. It could well be that functional relations exist between episodes in L1, in the sense that a problem which arises during episode A initiates a problem-solving sequence which is carried out in episode B. So, we can conclude that the writing process can be broken down into composing episodes in both languages. Furthermore, the only real differences between languages concerned the average number of episodes in each writing process and the variation in episode length.

Second, we found that, as expected, Planning did indeed occur more often in L1 than in L2. In addition, even though it was a very rare activity, it also occurred more often proportionally at episode boundaries than within episodes in both languages. But, contrary to our expectations, it did not turn out to be the most frequent activity at episode boundaries. Therefore the question remains, whether Planning has the ability to direct or influence the other activities that occur within

episodes. The other activities that were analysed also occurred more frequently, proportionally, at episode boundaries than within episodes. Furthermore, for two activities, Reading the assignment and Generating, we found an interaction effect: the difference between languages was larger at episode boundaries than within episodes.

Third, we predicted that, if activities such as Planning and Evaluating were able to guide or influence the activities that occur within episodes, then they would be positively correlated to text quality when they occurred at episode boundaries. Results indicated, however, that Planning was the only activity that correlated positively with text quality at episode boundaries. Flower and Hayes (1980a, 1981a) suggested that different types of Planning were hierarchically related. This holds to some extent in our study as well. We found that Planning at episode boundaries is more important for predicting text quality than Planning within episodes, which, in our view, is another argument for the existence of hierarchical relations between composing episodes within the writing process.

The occurrence of other potentially influential activities, such as Evaluating, at episode boundaries did not correlate with text quality at all. The only other activity that correlated with text quality at episode boundaries was Self-instructions, which appear to have a negative effect on text quality. The type of Self-instructions that are produced by weak writers are often comments such as *"let me see"* or *"let's read again"*, which might indicate that writers are stuck or have run out of ideas to write about. The occurrence of Self-instructions at episode boundaries could thus be an indication of a coping strategy that weaker writers fall back on when they find themselves in trouble during the writing process, rather than an indication of some form of (metacognitive) control over the writing process. Alternatively, it could be an unsuccessful attempt by weaker writers to exert metacognitive control over the writing process. However, whether the occurrence of Self-instructions is a sign of metacognitive control or merely a reflection of a coping strategy is probably a writer-dependent characteristic, possibly mediated by other factors, such as L1 writing skill and L2 proficiency. Which of the two it is, must be determined through qualitative analyses.

Finally, we were also interested in possible differences in Planning behaviour in L1 and L2. In line with earlier research (Jones & Tetroe, 1987) we found that writers were less likely to plan in L2 than in L1. But the overall effectiveness of their Planning behaviour was similar in both languages. On average, good writers planned twice as much as weaker writers at episode boundaries in L1 and three times as much at episode boundaries in L2. Planning is clearly the only activity that correlates positively with text quality when it occurs at episode boundaries. Thus we can conclude that Planning can influence the orchestration of the writing process. The writing process appears to consist of multiple concentration bursts, or composing episodes. When such episodes are

initiated by a Planning activity, this appears to have a positive effect on the quality of the resulting text, both in L1 and in L2, although the effect appears stronger in L1 than in L2. We found large effects for the overall differences between the occurrence of each activity at episode boundaries versus within episodes and medium effects for the differences between languages (for Planning) and the interaction effect between episode boundaries and languages (for Reading the assignment and Generating).

In earlier studies, we have consistently stressed the importance of including time as a factor in writing process research (see for example, Rijlaarsdam & Van den Bergh, 2006; Roca de Larios, et al., 2002; Chapter 3) and we analysed the moment at which activities occur during the writing process as a whole (see for example Chapter 3). In this study, we did not include time in the same way. Instead, we focused on the moment at which activities occur within composing episodes. Results indicated that it matters whether activities such as Planning and Self-instructions occur at episode boundaries or within episodes. So although we implemented a different operationalization of time in this study, our results confirm the importance of analyzing the moment at which activities occur during the writing process, albeit on another level.

Theoretical validity and psychological reality

In this study, we observed writers during on-line text production. Results clearly showed that they work in units of concentration or bursts of text production. The presence of composing episodes in the think-aloud protocols appears to mirror the presence of such units of concentration during the writing process. Therefore, dividing the writing process into composing episodes, based on indications in the data, appears to do justice to the way in which the writing process unfolds. Moreover, this suggests that the occurrence of episodes during the writing process is psychologically real. Indeed, analysing the writing process using composing episodes as the unit of analysis appears, in our view, to increase the psychological reality of descriptions of the writing process to date. The fact that it appeared possible to divide think-aloud protocols into composing episodes confirms the psychological reality and necessity of using a larger unit of analysis in writing process research.

Results of this study clearly indicate that the boundaries between composing episodes play a special role during the writing process. Cognitive activities, such as Planning, only appear able to influence text quality when they occur at episode boundaries. In other words, our idea that composing episodes occur during the writing process, and more importantly, that the presence of cognitive activities at episode boundaries is important for the quality of the resulting texts, seems theoretically valid. According to Johnson (1997): *"... theoretical validity is obtained to the degree that a theory or theoretical explanation developed from a research*

study fits the data and is, therefore, credible and defensible." (Johnson, 1997, p. 282). We hypothesized that Planning, though it is an infrequent activity, would be able to influence text quality indirectly, through its influence on the orchestration of the writing process as a whole. Our results indicated that this was indeed the case; Planning was the only activity that appeared to have a positive effect on text quality when it occurred at episode boundaries. Therefore, we would argue that our explanation for the way in which Planning influences text quality is theoretically valid.

Further research

In this study, we attempted to determine whether certain cognitive activities can exert influence over the orchestration of the writing process. Results indicate that Planning is the only activity which appears to influence text quality positively when it occurs at episode boundaries, both in L1 and in L2. Self-instructions appear to have a negative effect on text quality when they occur at episode boundaries in L1, but not in L2. So Planning at episode boundaries clearly appears beneficial for quality of the resulting texts. Still, causal relations between the occurrence of Planning and text quality cannot be inferred directly from these data. Experimental research must be carried out in order to confirm whether these relations are in fact causal ones. An experiment in which writers are stimulated to plan or prevented from Planning their writing in the beginning, the middle or near the end of the writing process might shed more light on this causality issue.

Finally, it is important to note that the design of this study consisted of a single task per participant per language. Earlier studies have shown that it is better to include multiple tasks per writer per condition for various reasons (see for example Chapter 2). Therefore the results of this study must be interpreted with caution. The results indicate that composing episodes and Planning at episode boundaries play an important role in the orchestration of the writing process. But these findings can only be seen as a first promising exploration of this theory. Further research, with multiple tasks per participant, will be carried out in order to confirm these results.

APPENDIX A: Example of an assignment

"Compulsory automatic donor registration": good or bad idea?
The NSU, the National Student Union, is organising a national essay contest, especially for students. You're also taking part. You absolutely want to win. The winning essay will be printed in all the university newspapers, including the U-blad. The U-blad is read by students and employees of the university.

The subject of the essay has already been decided and was described in the U-blad as follows:
Some politicians have recently proposed a new system for bringing in more organ donors: "compulsory automatic donor registration". In such a system, everybody 18 years or older automatically becomes a donor after death, unless they explicitly registered that they didn't want to. It is expected that the new system will make more organs available for transplantation. But not everybody thinks that "compulsory automatic donor registration" is a good idea. What do you think?

Assignment:
Write an essay in which you give your opinion on the question:

> "Everybody should automatically be a donor, unless they explicitly register against it: good or bad idea?"

The essay has to meet the following requirements, set by the Jury:

1. Your essay must be (about) half a page in length.
2. You must do your best to convince your readers, readers of university magazines, of your opinion.
3. You must give arguments to support your opinion.
4. Your essay must be structured in a good and logical way.
5. Your essay must look well-cared-for (think of language use and spelling).
6. In your essay you must use at least two extracts from the 'References' (see next page). You must include these extracts in your essay in a meaningful way.

You have 30 minutes to complete this assignment.

Good luck!

References

Every year hundreds of people die because they had to wait for an organ for too long. For example, in the Netherlands more than 1200 people are on the waiting list for a kidney transplant. This amounts to a waiting time of four to five years. The introduction of a system of "compulsory automatic donor registration" would be an efficient way of reducing these waiting times.
Adapted from: www.zorgkrant.nl, March 2005.

Organ donation is a form of life saving action. If you see someone drowning, you are obliged to help. In that case, choice is not an option: If you don't choose, you remain inactive, which is punishable by law. Donating organs can help and sometimes save people who are ill. This form of life saving action should be compulsory for everybody. Choosing not to be a donor, is failing to perform a life saving action.
Adapted from: N. Hoebe, letter to the editor, Trouw, 12th of March 2005.

Of more than eight million adults in the Netherlands, it is unknown whether or not they want to be a donor when they die. Among these eight million there are sure to be many people who would want to be a donor, but have simply never registered as such, for instance because they forgot. In a system of "compulsory automatic donor registration" the organs of all these people would suddenly become available. That's why the introduction of "compulsory automatic donor registration" is a good idea.
Adapted from: Agnes Kant, member of the Lower House for the Socialist Party, www.zorgkrant.nl, March 2005.

The more organ donors, the better. But always out of free will! Compulsory automatic donorship, unless you explicitly declare yourself against that, is not a good thing. "Silence is consent" is not a proper way of reasoning. It is a clever way of abusing the lack of knowledge of many people. Suddenly many more organs will become available, ripped from ignorant civilians, such as the homeless, who were not aware of becoming a donor after they died. And all of this under the pretext of "silence is consent".
Adapted from: 'Dave95' on the internet forum on www.debatplaats.vara.nl, March 2005.

"Until what age do my organs qualify for donation? How healthy do I need to be? Under what circumstances do I need to die? And most important of all: what is brain dead? It would be highly unpleasant if, during a near-death experience, I would see my organs being removed from my body! These are the questions that nobody is able to answer for us, especially not the Minister of Welfare and Health. Until then my answer will be 'no'."
Source: C. Putter, letter to the editor, Trouw, 12th of March 2005.

"How often do we have to say that people don't die because they didn't receive an organ, but because they were critically ill. In my view, "compulsory automatic donor registration" is undesirable."
Adapted from: Th. van der Kraats, letter to the editor, Trouw, 12th of March 2005.

CHAPTER 6

L1 USE DURING L2 WRITING: AN EMPIRICAL STUDY OF A COMPLEX PHENOMENON

Abstract [1]

Earlier research has generated conflicting results regarding the way in which second language (L2) writers use their first language (L1) when writing in their L2. Therefore, this study examines when writers use their L1 to carry out conceptual activities. The focus is on when L1 use occurs, whether it varies between writers and tasks, and whether L1 use is related to L1 writing skill, L2 proficiency and text quality. To answer these questions, twenty students each wrote four short argumentative essays in L1 (Dutch) and four in L2 (English) under think-aloud conditions. Results indicate that participants all use their L1 while writing in L2 to some extent, although this varies between activities. In addition, the extent to which L1 is used to carry out different activities appears to vary somewhat between tasks. Furthermore, the amount of L1 use is negatively related to both L1 writing skill and L2 proficiency, for most but not all activities. Finally, L1 use is negatively correlated to L2 text quality. This relationship appears to be influenced most strongly by L1 writing skill, but also by L2 proficiency, as L1 use appears to have a negative influence on the quality of texts produced by writers with low L1 writing skill and by writers with high L2 proficiency. This implies that L1 use does not necessarily have to be discouraged for all writers or under all circumstances. Overall, L1 use during L2 writing seems to be a rather complex phenomenon.

Keywords: Generating ideas; L1; L2; Metacomments; Planning; Language switching; Text quality; Writing process;

Earlier research has shown that writers use their first language (L1) while writing in L2, although the extent to which they do so clearly varies (see for example, Krapels, 1990; Friedlander, 1990; Uzawa, 1996; Woodall, 2002). Studies to date have found that L1 is used by adult writers during L2 writing for a wide variety of purposes, such as *planning* (Beare, 2000; Jones & Tetroe, 1987; Krapels, 1990; Uzawa & Cumming, 1989; Wang, 2003; Woodall, 2002), *generating ideas or content* (Beare, 2000; Beare & Bourdages, 2007; Knutson, 2006; Krapels, 1990; Qi, 1998; Roca de Larios, Murphy, & Manchón, 1999; Uzawa & Cumming, 1989; Wang, 2003; Whalen & Ménard, 1995; Woodall, 2002), or for *solving linguistic problems* such as vocabulary

---

[1] Van Weijen, D., Van den Bergh, H., Rijlaarsdam, G., & Sanders, T. (2008). *L1 use during L2 writing: An empirical study of a complex phenomenon*. Manuscript submitted for publication.

issues (Beare, 2000; Cumming, 1989; Jones & Tetroe, 1987; Lay, 1982; Wang, 2003; Woodall, 2002). Furthermore, L1 use has also been reported for back-tracking (Manchón, Roca de Larios, & Murphy, 2000), stylistic choices (Knutson, 2006), and as a means to prevent cognitive overload (Cohen & Brooks-Carson, 2001; Knutson, 2006; Qi, 1998; Woodall, 2002).

Studies which reported on L1 use, such as those above, were carried out for a number of different reasons and with varying research goals. First of all, some studies focused on comparisons of L1 and L2 writing (e.g., Uzawa & Cumming, 1989; Whalen & Ménard, 1995) or on the extent to which writers transfer their L1 strategies to L2 writing (e.g., Wolfersberger, 2003). These studies see L1 use as a strategy which writers employ during L2 writing, mainly in the form of translating from L1 to L2. This view is also shared by other studies, which focused on the influence of learner characteristics such as writing expertise and L2 proficiency on L2 writing (e.g., Cumming, 1989; Sasaki, 2002, 2004; Sasaki & Hirose, 1996). Second, there were several studies which included L1 use as an independent variable, for example by instructing participants to plan in L1 or L2 before writing their L2 texts (Akyel, 1994; Friedlander, 1990; Lally, 2000) or by instructing participants to write a text in L1 and then translate it into L2 (e.g., Cohen & Brooks-Carson, 2001; Kobayashi & Rinnert, 1992). Comparisons between the translate and direct writing (L2-only) conditions were complicated, however, by the fact that participants in the direct writing condition reported using L1 very often while writing in their L2, even though they were not supposed to (Cohen & Brooks-Carson, 2001, p. 179; Kobayashi & Rinnert, 1992, p. 200).

Third, an increasing number of studies focused specifically on the role L1 use plays during L2 writing. These ranged from small-scale case studies, which were mainly exploratory in nature (e.g., Knutson, 2006; Lay, 1982; Qi, 1998), to studies which attempted to determine how L1 use is related to specific activities, such as *generating ideas* (Beare & Bourdages, 2007), *backtracking* (Manchón, et al., 2000), or *planning* (Jones & Tetroe, 1987). Finally, some studies investigated the effect of L2 proficiency on L1 use (e.g., Wang, 2003; Wang & Wen, 2002; Woodall, 2002). Some of these studies, however, were rather vague about what L1 use actually means and how it was measured. Lay (1982) for example, found more L1 use on certain topics than on others and reported that more L1 use resulted in better quality texts. But it is unclear what *"more L1 use"* actually means (see Lay, 1982, p. 406). Other studies were more precise and attempted to calculate to what extent L1 was used during writing in L2, by reporting: the overall percentage of L1 words in L2 think-aloud protocols (Wang & Wen, 2002), the mean number of language switches per task (Wang, 2003; Woodall, 2002), and the length of time (number of seconds) that L1 use occurred during L2 writing (Woodall, 2002).

Despite their different research goals, many of the studies described above attempted to relate L1 use to either L2 proficiency or text quality in some way. But,

not surprisingly, given their different goals, they produced rather conflicting results. On the one hand, several studies reported that high proficiency writers switched more between L1 and L2 than low proficiency writers (see for example, Wang, 2003). In addition, Cumming (1989) concluded that expert writers used L1 frequently during word searches. On the other hand, there are studies which concluded more or less the opposite. According to Sasaki & Hirose (1996), weak writers reported translating more from L1 to L2 while writing than good writers, although the difference between the two groups failed to reach significance. In two subsequent studies, Sasaki (2002, 2004) found that novice writers translate more often from L1 to L2 than expert writers, and that novices also continue to do so over time (Sasaki 2004). Similarly, Wang & Wen (2002) concluded that lower proficiency writers used L1 far more than higher proficiency writers. Wolfersberger (2003), who only studied low proficiency L2 writers, also found that they frequently use L1 during pre-writing and make use of translating from L1 to L2 in order to compensate for their limited ability to write in L2. In line with this, Beare and Bourdages (2007) found that high proficient bilingual writers hardly used L1 at all during L2 writing (Beare & Bourdages, 2007, p. 159). Moreover, Woodall (2002) complicated the discussion even further, by including the difference between cognate and non-cognate languages as an additional variable in his study. He found that overall, intermediate proficiency writers switched more often from L1 to L2 than high proficiency writers, but this effect was influenced by whether they were writing in non-cognate (Japanese/English) or cognate languages (Spanish/English). Therefore, Woodall concludes that there seem to be important differences in L1 use between writers. Indeed: *"some students appeared to control their L-S [language switching], using their L1 as a tool. For others, L-S seemed out of control, and the L1 seemed more like a crutch to obtain cognitive stability"* (Woodall, 2002, p. 20).

Unfortunately, studies directly relating L1 use during L2 writing to text quality are few and far between, but there are indications that both translation from L1 to L2 and L1 use during L2 writing can be beneficial for some writers (Cohen & Brooks-Carson, 2001; Kobayashi & Rinnert, 1992; Uzawa, 1996; Uzawa & Cumming, 1989). In addition, Knutson (2006) concluded that L1 use does not always have a negative effect on text quality, because her second best writer used L1 regularly, but still produced good texts. Woodall (2002) also seemed surprised by the fact that L1 use appeared to be positive for high proficiency writers of cognate languages (Woodall, 2002, p. 20).

Finally, some studies focused on the possible effect of task features, such as topic-knowledge or cultural factors, on L1 use during L2 writing and text quality, but they all failed to find a significant effect of planning during pre-writing in L1 or L2 on text quality (Akyel, 1994; Friedlander, 1990; Lally, 2000). However, Friedlander (1990) did find that writers wrote their best texts on familiar topics related to their L1 cultural background, regardless of whether the plans for those

texts were produced in L1 or L2. Krapels (1990) and Lay (1982) also concluded that tasks on L1-related topics generated more L1 use during L2 writing than other tasks. But the effect of this topic-knowledge factor on text quality remains somewhat unclear (Qi, 1998).

Although the studies reviewed above have provided new insights into the possible role L1 plays during writing in L2, overall, a somewhat foggy picture emerges. The use of L1 during L2 writing can be beneficial, but not in all cases and not for all writers (Cohen & Brooks-Carson, 2001). This appears to depend on L2 proficiency (Akyel, 1994; Beare & Bourdages, 2007; Wang, 2003; Wang & Wen, 2002; Wolfersberger, 2003; Woodall, 2002), the type of task (Wang & Wen, 2002), topic-knowledge (Friedlander, 1990, Krapels, 1990, Lay, 1982; Qi, 1998), or on whether L1 and L2 are cognate or non-cognate languages (Woodall, 2002). Furthermore, the reasons for L1 use and the cognitive activities that are carried out in L1 also remain somewhat unclear. L1 can be used to solve linguistic or lower-order problems (Beare, 2000; Jones & Tetroe, 1987; Qi, 1998; Wang, 2003; Woodall, 2002), but it is also used for higher-order activities, such as planning, or to prevent cognitive overload (Beare 2000; Cohen & Brooks-Carson, 2001; Jones & Tetroe, 1987; Knutson, 2006; Krapels, 1990; Qi, 1998; Roca de Larios, et al., 1999; Uzawa & Cumming, 1989; Wang, 2003; Whalen & Ménard, 1995; Woodall, 2002).

The main reason for these conflicting results is probably the fact that most of the studies discussed above were rather small-scale, qualitative studies, with varying L1's and methods, which were based on a rather small number of participants, usually eight or less (Beare, 2000; Beare & Bourdages, 2007; Jones & Tetroe, 1987; Knutson, 2006; Manchón, et al., 2000; Qi, 1998; Roca de Larios, et al., 1999; Uzawa & Cumming, 1989; Wang, 2003; Wolfersberger, 2003). But there are also a number of other methodological problems. First of all, a single-task-per-condition design was often used, which means that participants were required to write a single text in each condition (Akyel, 1994; Beare, 2000; Beare & Bourdages, 2007; Friedlander, 1990; Manchón, et al., 2000; Uzawa & Cumming, 1989; Wang, 2003; Wang & Wen, 2002; Whalén & Ménard, 1995; Wolfersberger, 2003; Woodall, 2002). This is problematic, because research has indicated that both writing processes and text quality can vary strongly within writers (see for example, Schoonen, 2005; see also Chapters 2 & 4). This means that the extent to which L1 use occurs and the influence of L1 use on text quality cannot be assessed reliably based on a single text per writer.

A second problem is the fact that it is unclear how L1 use is related to the occurrence of different cognitive activities, such as generating ideas and planning during L2 writing. Previous research has indicated that both planning and generating appear to have a positive effect on text quality in L2 writing, depending on the moment at which they occur during the writing process (see Chapters 3 &

5). But the positive effect of such activities on text quality might be mediated by the language in which they occur. The question, then, is whether the occurrence of each activity in L1 while writing in L2 also has a positive effect on L2 text quality. As a result, a distinction must be made, for example, between L1 and L2 planning during L2 writing, so that the effect of L1 use on the relationship between cognitive activities and text quality can be determined.

A third and final problem is related to the way in which L1 use was analysed in earlier studies. Many studies did not provide a clear definition or measure of L1 use because they were small-scale, exploratory studies (e.g., Knutson, 2006; Lay, 1982; Qi, 1998), or because L1 use was not the main focus of the analysis (Cumming, 1989; Sasaki, 2002, 2004; Uzawa & Cumming, 1989; Whalen & Ménard, 1995; Wolfersberger, 2003). Studies that did attempt to measure L1 use more accurately, either included writers' self-reported percentage of L1 use (e.g., Cohen & Brooks-Carson, 2001; Kobayashi & Rinnert, 1992; Sasaki & Hirose, 1996) or reported the duration or length of L1 use as a proportion of the writing process as a whole (Woodall, 2002). This is questionable, because it does not provide any information on the occurrence of different cognitive activities in L1. A few studies did investigate the proportion of L1 use for specific cognitive activities, albeit in slightly different ways (Beare, 2000; Beare & Bourdages, 2007; Jones & Tetroe, 1987; Manchón, et al., 2000; Wang & Wen, 2002). Furthermore, there are indications that not all activities occur in L1 during L2 writing to the same extent. Wang and Wen (2002) found that conceptual activities, such as planning and generating ideas are more likely to occur in L1 during L2 writing than linguistic activities (see Wang & Wen, 2002, p. 239). This might be related to the fact that conceptual activities such as planning generally require more cognitive effort than linguistic activities such as formulating (see Kellogg, 1994, p. 17; Rijlaarsdam, Braaksma, Couzijn, Janssen, Kieft, Broekkamp, & Van den Bergh, 2005, p. 129; see also Stevenson, 2005). If writers experience cognitive overload due to the increase in task complexity that writing in L2 often entails, then it seems likely that they will revert to L1 use for the most demanding activities, in this case the conceptual ones. But again, the studies on L1 use mentioned above did not relate their findings to text quality directly, so the influence of the language in which different activities occur on text quality remains unclear.

In sum, these problems combined make it hard to generalize results over tasks or across languages, especially considering the fact that L1 use appears to differ between cognate and non-cognate languages (see Woodall, 2002). Moreover, many studies attempted to relate L2-proficiency to L1 use (Beare, 2000; Beare & Bourdages, 2007; Uzawa & Cumming, 1989; Wang, 2003; Wang & Wen, 2002; Whalen & Ménard, 1995; Wolfersberger, 2003), instead of examining the possible effect of L1 use on text quality directly. Although determining a direct link between L1 use and text quality may be difficult, it does seem potentially far more

informative, especially for educational purposes. Therefore, we hope to provide further insight into the role L1 use plays during L2 writing, by carrying out an analysis based on multiple tasks per writer. In addition, it seems advisable to focus on specific activities and then determine the extent to which each activity occurs in L1, as a proportion of how often each activity occurs during the writing process as a whole. By relating the proportion of L1 use for each activity to text quality, we can hopefully determine the influence L1 use has on text quality more accurately.

Aims and research questions

The aims of this study are twofold. First, we wish to determine to what extent specific cognitive activities occur in L1 during L2 writing. In line with earlier research, we will focus on conceptual activities, as these are more likely to occur in L1 than linguistic activities (Wang & Wen, 2002, p. 239 – 240). By asking participants to write more texts each, we hope to determine whether L1 use is relatively stable within writers or whether it varies between tasks. This can also help us determine whether writers are more or less inclined to use L1 while writing in L2 due to personal preferences or whether this is influenced by learner characteristics, such as L1 writing skill or L2 proficiency, or by task effects such as the topic of the assignment (see Krapels, 1990, p. 53). Second, we would like to establish more directly what the effect of L1 use is on text quality and to what extent this is influenced by the learner characteristics: L1 writing skill and L2 proficiency. This will be done by analyzing which cognitive activities occur in L1 during L2 writing and whether their occurrence correlates with text quality. Therefore, the main questions in this exploratory study are: *(1) Do writers use their L1 to carry out conceptual activities while writing in L2? And if so, when?* and *(2) Is L1 use while writing in L2 related to text quality?*

In order to answer the first question, we must determine whether the amount of L1 use differs between activities, whether the amount of L1 use is stable within writers, and whether the amount of L1 use is related to L1 writing skill and/or L2 proficiency. With regard to the second question, we must determine whether L1 use correlates directly with observed text quality scores, and whether the correlation is influenced by L1 writing skill and/or L2 proficiency. Finally, the importance of including an operationalization of the changing task representation in writing process research was consistently stressed in earlier studies (see for example, Breetvelt, Van den Bergh, & Rijlaarsdam, 1994; Rijlaarsdam & Van den Bergh, 2006; Van den Bergh & Rijlaarsdam, 2001; Van den Bergh, Rijlaarsdam, Janssen, Braaksma, Van Weijen, & Tillema, in press). Therefore, in this study, we will also attempt to determine when writers use their L1 during L2 writing.

Regarding question one, we predict that L1 will be used by most of the writers during writing in L2 (see also Wang & Wen, 2002). However, the extent to which they use L1 when they write will probably vary between writers. We also

predict that the extent to which different conceptual activities occur in L1 during L2 writing will be related. This means that if planning occurs frequently during L2 writing, then activities such as metacomments or self-instructions will as well. In addition, we predict that L1 is probably a writer-specific characteristic. This means that for each writer the proportion of L1 use will be rather stable or similar between tasks. Furthermore, we expect both L2 proficiency and L1 writing skill to have a negative effect on L1 use; skilled L1 writers and writers with high L2 proficiency will be less likely to use L1 while writing in L2 than weaker writers. Regarding question two, we predict that L1 use will be related to text quality, although the correlations may vary between activities. However, due to the conflicting results of earlier studies, which state that L1 use may or may not have a negative effect on text quality, we cannot accurately predict whether L1 use will have a positive or a detrimental effect on text quality. We do predict, however, that both L2 proficiency and L1 writing skill will influence the correlation. In other words, the effect of L1 use on text quality will differ for writers with high or low L2 proficiency or L1 writing skill.

METHOD

Design

In this study, twenty participants were asked to write four texts each, under think-aloud conditions. Participants were all first-year BA English majors, whose L1 was Dutch. The majority was female (85%), which was in line with the male-female ratio of the students in the Arts faculty. The study took place during their first semester at university, and their average age was 18 years. The students were all paid a small amount for their participation in the study.

The set-up of the study required each student to write four short essays in their L1 (Dutch) and four short essays in English (L2). Each participant attended four individual sessions, in which they wrote two texts on the computer, under think-aloud conditions. Each writing process was recorded with audio and video equipment, so that think-aloud protocols could be produced. They were given 30 minutes to write each text, but they were free to stop when they thought they were finished, and were given a few extra minutes to round off their texts if necessary. Therefore, the average time they spent writing each essay was not 30 minutes, but 23.46 minutes (*sd* = 7.23 minutes) for L1 and 26.10 minutes (*sd* = 6.43 minutes) for L2. Time on task also varied enormously between tasks in both languages (L1: 10.79 to 37.64 minutes; L2: 12.83 to 36.89 minutes)

Participants were trained in thinking aloud during their first individual session. The training consisted of solving an algebra puzzle, a small crossword puzzle and writing a short text of about five lines while thinking aloud. These short

texts had to be written on one of ten randomly assigned topics, such as describing their favourite music, book, film or what they did last weekend. Such a text had to be written on one of the other topics at the start of each subsequent writing session, to reactivate their think-aloud mode. Finally, participants were told that they were free to think-aloud in L1, L2 or both while writing their texts (see also Knutson, 2006, p. 97).

The use of thinking aloud as a tool in writing research has been much debated in the past (see for example, Ericsson, 1998; Janssen, Van Waes, & Van den Bergh, 1996; Manchón, Murphy, & Roca de Larios, 2005; Roca de Larios, Manchón, & Murphy, 2006). However, we agree with Krapels (1990) that protocol analysis provides *"more useful data"* for L2 writing research (Krapels, 1990, p. 51). Furthermore, we decided to use the method in this study, because it is, in our view, one of the best methods for observing the occurrence of conceptual activities such as planning, generating ideas and evaluating, during the writing process. In the past, some studies required writers to report to what extent they thought they used L1 while writing in L2 (e.g., Cohen & Brooks-Carson, 2001; Kobayashi & Rinnert, 1992), but if we want to determine which activities actually take place in L1 during L2 writing, and more importantly when L1 use occurs, protocol analysis seems the only appropriate method available. We chose to collect concurrent protocols instead of retrospective protocols, because concurrent verbalizations seem most directly related to what writers think while writing. Furthermore, concurrent verbalizations seem more valid than retrospective protocols: *"because of the absence of a time interval between the occurrence of a thought and its expression"* (Roca de Larios, et al., 2006, p. 104; see also Ericsson, 1998; Hayes & Gradwohl Nash, 1996, p. 45). Therefore, we felt that thinking aloud was the most appropriate research tool for this study.

Instruments

The assignment was to write a short essay on one of eight topics. The basic assignment was the same in all cases, only the topic varied between tasks. Each time participants had to write a short essay for a made-up essay contest in the local University Newspaper. The assignment described the goal, the topic and the target audience of the essay. It also contained six short sources related to the topic, of which they had to incorporate at least two in their texts. The essay had to include the writer's own opinion, supported by arguments for or against the topic. Each participant was assigned four topics to write on in each language, based on their participant number. The odd numbered participants (1 through 19) wrote essays in English on: *Compulsory Organ Donation, Downloading Music*, *Life as a Student,* and *Soft Drug Legislation*, while the even numbered participants (2 through 20) wrote in English on the other four topics: *Education in English*, *Surveillance Cameras*, *Having Children,* and *Mobile Phone Use*. Each group wrote on the four other topics in L1

Dutch. The L1 essays were collected to determine participants' L1 writing skill. An example of an assignment is included in Appendix A. Participants wrote their essays in random order, to minimize the risk of task-order or language effects.

Table 1: Overview of the eleven main categories included in the coding scheme

Cognitive activity	*Description and protocol example*
(Re)-Reading the assignment	Reading the assignment and/or the sources
Planning: Self-instructions	Instructions the participant gives himself regarding the next step in the writing process (e.g., *"I need to decide how to start"*)
Planning: Goal setting	Paraphrasing task demands in the assignment or setting new goals (e.g., *"This text is going to be about..."*)
Planning: Structuring	Selecting and ordering ideas or making a rough outline (e.g., *"First the pro's, then the con's and then my own opinion!"*)
Generating ideas	Paraphrasing ideas from the assignment or generating new ones: (e.g., *"Us humans we're very curious and we like to gossip!"*)
Metacomments	Reflections on the writing process as a whole or comments on the assignment and sources (e.g., *"I had the perfect sentence in my head"*)
Pausing	Silence or interjections (e.g., *um, ehhh*)
Formulating	Text production or typing the actual words on the computer.
Rereading own text	Rereading segments of the text produced so far
Evaluating own text	Evaluating the text produced so far at word, sentence or text level (e.g., *"No I'm not gonna say this"*)
Revising own text	Revising the text produced so far at word, sentence or text level (e.g., erases 'them' and types 'people')

During the last session, participants were asked to complete a vocabulary test, in order to obtain a measure of their L2 proficiency. The test consisted of 64 items, which were ordered from easy to difficult, and its internal consistency was good (α = .81). Each item was an English sentence containing a blank, which participants had to fill in. The Dutch translation of the word in the blank was provided to help them. Thus it was both a vocabulary and a translation exercise. Participants were given seven minutes to complete as many items as they could and they earned a point for every item that they completed correctly. Because it was a timed test, this enabled us to determine how fast and how accurately participants were able to find a suitable translation for an L1 word within a certain context. Scores ranged from

48% to 92%, with an average of 72% (*sd* = 10.4), which indicates that the test's level of difficulty was appropriate for this group of students. Finally, participants were divided into two groups with high and low L2 proficiency based on their test scores.

Data

The data in this study were the writers' 79 L2 think-aloud protocols[2], which were produced by transcribing the audio and video recordings of each L2 writing session. Each protocol was divided into segments, based on the writer's behaviour. If a writer switched from one activity to another (e.g., from Planning to Formulating), then the segment containing the first activity ended and the new activity was placed in the next segment (see Breuker, Elshout, Van Someren, & Wielinga, 1986). The protocol segments were subsequently coded using a coding scheme based on Hayes and Flower's (1980) model of the writing process. The coding scheme was used regularly in earlier research (see for example, Breetvelt, et al., 1994; Couzijn, Van den Bergh, & Rijlaarsdam, 2002; Van den Bergh & Rijlaarsdam, 1999, 2001). The main categories in the coding scheme are shown in Table 1. The coding was carried out by six different coders, who each coded 16 protocols. Each coder coded four protocols which overlapped with one or more other coders, so that the interrater reliability could be determined, which was high (Cohen's kappa = 0.95).

Text quality

The quality of the L2 texts (n = 79) was assessed by a group of five raters in two different ways: analytically and holistically. The texts were first rated analytically, using a rating scheme with four categories: *Structure, Content, Argumentation* and *Conclusion*. The raters performed the holistic rating two weeks later. For this rating, raters received a benchmark essay, worth 100 points, for each topic. Their task was to decide how much better or worse the quality of each essay was than the benchmark essay. The reliability between raters was sufficient for both methods (analytical: α = .93; holistic: α = .83) and the correlation between the two ratings was relatively high (r = .69; corrected for attenuation, r =.79). Therefore, the scores of the two ratings were combined to form a single text quality score for each essay. The averages scores per topic and the overall average score are presented as z-scores in Table 2. The mean scores for some topics are higher than for others, but text quality was not affected by the order the texts were written in (F (3, 75) = .10; p = .96).

As mentioned earlier, the L1 texts (n = 80) were collected to determine writers' L1 writing skill. The quality of these texts was assessed in the same way as the L2 texts, but by a different group of five raters. The between rater reliability

[2] One participant wrote three essays in L2 instead of four due to technical difficulties.

was satisfactory for both the analytical (α = .88) and the holistic method (α = .82) and the correlation between the two ratings was also rather high (*r* = .74, or .87 when corrected for attenuation). Hence, a single z-score was calculated for each essay. Finally, L1 writing skill was calculated as the average score for L1 text quality for each writer (mean: -.01, range: -1.99 to 1.26; *sd* = .98). Participants were then divided into two groups with high and low L1 writing skill based on their scores.

Table 2: Average text quality and range per topic in z-scores (N = 79)

Task	*Topic*	*N*	*Mean (sd)*	*Min.*	*Max.*
A	*Education in English: good idea or bad suggestion?*	9	-0.15 (1.31)	-1.96	1.72
B	*Having children, yes or no?*	10	0.17 (0.97)	-1.32	1.68
C	*Everybody should automatically be a donor, unless they explicitly register against it: good or bad idea?*	10	0.19 (1.09)	-1.49	1.82
D	*Life as a student: A hard life full of stress? Or isn't it that bad?*	10	0.31 (0.45)	-0.55	0.80
E	*Do surveillance cameras in inner city areas increase public security?*	10	0.15 (1.23)	-2.36	1.54
F	*Mobile phones: irritating or essential?*	10	-0.32 (1.34)	-3.08	1.53
G	*Downloading music for free: criminal or should be possible?*	10	-0.17 (0.53)	-1.37	0.66
H	*Soft drugs: should they be legalised or prohibited completely?*	10	0.18 (0.89)	-1.46	1.77
Total		79	0.05 (1)	-3.08	1.82

Analysis

Research by Wang and Wen (2002) suggests that L1 use depends on the extent to which activities are related to the text itself: *"the more the cognitive processing is related to the textual output, the less L1 is used in it."* (p. 239 – 240). So, conversely, L1 is most likely to be used for the activities which are least directly related to the textual output. Therefore, the analysis focused on a number of conceptual activities, for which L1 seems most likely to be used while writing in L2: *Generating ideas, Metacomments,* and *Planning,* which is sub-divided into three different activities: *Self-instructions, Goal setting,* and *Structuring* (see Table 1). Generating ideas is not considered to be a type of Planning, because it is more closely related to the content of the text. Metacomments were defined as comments which were not directly related to the content of the text, but more to the way the writing process

was carried out or to more general aspects of the writing process as a whole (see also Jones & Tetroe, 1987, p. 49). In line with earlier research, we will focus on the *"non-prescribed use of the L1 in L2 writing"* (Woodall, 2002, p. 8). Each protocol segment which contains one or more L1 words is considered an occurrence of L1 use. In addition, the analysis will focus on the proportion of L1 use for each cognitive activity. In other words, for each cognitive activity we will determine the proportion of protocol segments that contain L1 use. This proportion will then be related to L2 proficiency, L1 writing skill and to text quality, in order to answer the research questions. Finally, we will also determine when writers are most likely to use their L1 while writing in L2.

RESULTS

L1 use in L2 writing

Results indicate that every writer in this study used L1 at least once while writing in L2, but some writers used L1 far more frequently than others. Subsequently, we carried out a multilevel analysis, to determine when activities are most likely to be carried out in L1 while writing in L2. Results indicate that each of the five activities was slightly more likely to occur in L1 at the start of the writing process than near the end. Overall, the probability of occurrence for L1 use decreased slightly over time from .11 in the beginning to .09 near the end of the writing process, but this was only a small effect. Furthermore, we found no significant variation between writers or tasks except in the case of Generating. Thus, whether writers Generate ideas often in L1 or not seems to be slightly writer dependent.

Table 3: Average proportion of L1 use for each activity

Cognitive activity	*N (overall)*	*Average L1 use (sd)*	*Min.*	*Max.*
Self-instructions	1827	0.45 (0.39)	0.00	1.00
Goal setting	60	0.07 (0.25)	0.00	1.00
Structuring	95	0.19 (0.39)	0.00	0.99
Generating ideas	3272	0.14 (0.20)	0.00	0.66
Metacomments	801	0.43 (0.39)	0.00	1.00

The next step was to determine whether the proportion of L1 use differed between activities. Table 3 shows the proportion of L1 use for each activity, averaged over writers and tasks. This indicates that, on average, the proportion of L1 use is less than 50% for all five activities. However, some activities, such as Self-instructions

(45%) and Metacomments (43%) are far more likely to occur in L1 while writing in L2 than other activities. In addition, the variation in L1 use is rather high for all activities; the percentage of L1 use over writers ranges from 0 to 100% for all activities, except Generating ideas, for which L1 use was never higher than 66%. This indicates that activities occur only in L2, only in L1, or in a mixture of L1 and L2 use while writing in L2. Furthermore, we carried out a repeated measures analysis in order to determine whether the differences in proportion of L1 use between activities were significant. The differences between activities were significant for all pairs of activities (F (1, 78) > 4.44; p <.04), except for two pairs: Generating and Structuring, and Self-instructions and Metacomments. No significant differences in the proportion of L1 use could be found for those two pairs, even though their overall frequencies of occurrence do differ (see N in Table 3).

Some examples of L1 use for each activity are presented in Table 4. These examples show that Self-instructions and Metacomments often occur in L1 when participants encounter a problem or appear to feel frustrated, while the other three activities appear to occur in more neutral contexts (see Table 4). Finally, it is important to note that Goal setting and Structuring, two types of Planning, occurred extremely infrequently overall during writing in L2, on average in less than 1% of protocol segments while writing in L2. In addition, we found that the proportion of L1 use for these activities correlated significantly with their overall proportion of occurrence during the writing process (Goal setting: r = .51, p <.001; Structuring: r = .41, p <.001). This suggests that writers who were likely to engage in Goal setting or Structuring, were likely to do so in L1 while writing in L2.

Variation in L1 use within writers

The next question was whether the occurrence of L1 use for each activity was stable within writers. To answer this question, two other questions must be answered first. The first question is: to what extent does the use of L1 while writing in L2 vary between writers? Are there participants who use L1 frequently and others who hardly use it at all? Therefore, we correlated the proportions of L1 use for each activity with one another, to determine whether the extent to which activities occur in L1 is related (see Table 5). If there is a strong correlation between the proportions of L1 use for two activities, then writers who carry out the first activity often in L1 while writing in L2 are likely to carry out the second activity in L1 as well. Table 5 shows that only Self-instructions and Metacomments share a high correlation (r = .75, p < .001), but they have 56% common variance. For all the other activities, the correlations are much lower, and the lowest, between Goal setting and Generating ideas is not even significant (r = .14, p = .21). This indicates that writers do not appear to execute their conceptual activities exclusively in L1 or L2 while writing in L2 for all tasks and all activities. However, if writers are likely to

carry out Self-instructions in L1 then they are also likely to do so for Metacomments.

Table 4: Examples of L1 use during L2 writing for each cognitive activity

Cognitive activity	*Protocol segment (participant, task)*	*Translation of L1 use*
Self-instructions	this means that ehm (..) everybody ehm becomes (.) that (.) oooh hoe druk je dat uit? this means that everybody (..) ehm (..) automatically is a donor (3, C)	oooh how do you say that?
	[Reads the assignment:] Everyone can imagine that classes in English (...) nou dan moet ik maar 'ns beginnen met wat schrijven dat lijkt me ook wel 'n idee (..) [Reads the assignment:] "Education in English: good idea or bad suggestion? (10 A)	well then I should start writing something, that sounds like a good idea
Goal setting	If nee (..) maar ik wil eigenlijk nog even samenvatten wat ik eigenlijk heb gezegd (...) (1, D)	no (..) but really I want to summarize what I've actually just said
	(..) dus dat is sowieso mijn standpunt (.) en dan moet ik de mensen een beetje gaan overtuigen daarvan [turns over the page] samen met deze (..) twee bronnen (..) [Reads the assignment:] you must use at least two extracts from the 'References' (see next page). (10 A)	so that is my point of view (.) and then I should try to convince people of it (..) together with these (..) two sources
Structuring	[Reads the assignment:] Piet Hein Donner, Minister of Justice, {...} doesn't feel anything for the proposal {...} According to the mayors, that would (.) decriminalize the weed cultivation, but Donner disputes that. (..) even kijken, misschien kan ik er wel twee argumenten van maken [splits paragraph into two paragraphs] uhm so the ministry of Justice (3, H)	maybe I can turn it into two (..) arguments

(table continues)

Table 4 (continued)

Cognitive activity	*Protocol segment (participant, task)*	*Translation of L1 use*
Generating ideas	I think it's essential [turns over the page] (.) uhmm (...) [researcher asks: what are you thinking?] PP: I was thinking of (.) ik ben ehh inleiding aan 't verzinnen ik dacht misschien kan ik een [laughs out loud] mobiel geluidje doen maar dat kan je niet typen ehmm (..) even kijken ehmm (..) [laughs out loud] (.) ehmm (..) [then rereads the assignment] (6, F)	I'm uhm trying to come up with an introduction, I thought maybe I could {laughs out loud} do a mobile phone noise but I can't type that let me see
Metacomments	which is recognized by (..) beetje valsspelen maar ja (.) anders heb ik echt geen bronnen [starts to reread the sources] (4, E)	it's cheating a little but well (.) otherwise I really won't have any sources [in my text]
	wants to have (...) ohhh! ik weet echt even niet wat ik wil zeggen (..) (8, B)	ohhhh! I really don't know what I want to say right now
	the network services are not totally responsible (.) for the decline of CD (.) sales (.) and (..) and (...) downloading music will (..) lead to satisfied customers (....) ok (.) ik spreek mezelf wel een beetje tegen (...) (9, G)	I'm sort of contradicting myself a little
	Bilingual (.) oh mag geen vraagtekens in de titels dat heb ik geleerd (..) dan gaan ze helemaal flippen (..) bilingual (..) [reading the assignment:] offers you the chance to lay your foundation (10, A)	oh I was taught that I can't use question marks in titles (..) then they'll totally freak out

Table 5: Correlations between the proportion of L1 use between cognitive activities

Cognitive activity	*SI*	*GS*	*St*	*GI*	*MC*
Self-instructions (SI)	---				
Goal setting (GS)	.26*	---			
Structuring (St)	.47**	.33**	---		
Generating ideas (GI)	.66**	.14	.44**	---	
Metacomments (MC)	.75**	.20	.38**	.64**	---

* ($p < .05$), ** ($p < .01$)

The second question is: how stable is L1 use over tasks? This is indicated by the standard deviation between tasks within writers. These standard deviations are averaged to get a more general picture. Results indicate that Generating is rather stable between tasks (mean *sd* = .08), while Goal setting (mean *sd* = .14), Metacomments (mean *sd* = .16), Self-instructions (mean *sd* = .17), and Structuring (mean *sd* = .23) show increasing signs of between task variation. Thus we can conclude that L1 use is more likely to vary between tasks for some activities than for others. Overall, we can say that the extent to which writers use their L1 while writing in L2 does not generally appear to be a writer specific characteristic, because the proportion of L1 use is not strongly correlated for all activities and because the extent to which the proportion of L1 use varies between tasks, varies more for some activities than for others.

The influence of L2 proficiency and L1 writing skill on L1 use

To determine whether L1 use is related to L2 proficiency, we correlated writers' scores on the vocabulary test with the proportion of L1 use for each activity. Significant negative correlations were observed between L2 proficiency and L1 use for Generating ($r = -.27$; $p < .05$), Metacomments ($r = -.29$; $p < .01$), and Self-instructions ($r = -.27$; $p < .05$). This means that lower L2 proficiency writers are more likely to use L1 for these activities than high L2 proficiency writers. The correlations for Goal setting ($r = -.045$; $p = .69$) and Structuring ($r = -.18$; $p = .12$) were not significant. Thus, L2 proficiency seems to influence L1 use, but not for all activities.

Subsequently, we correlated the proportion of L1 use for each activity with writers' average L1 text quality scores, to determine whether L1 use is related to L1 writing skill. Again, we found negative correlations for several, but not all activities. L1 use and writing skill appear to be negatively related for Goal setting ($r = -.22$; $p < .05$), Generating ($r = -.26$; $p < .05$), and Metacomments ($r = -.24$; $p < .05$). This means that L1 is more likely to be used to carry out these activities by writers with lower writing skill in L1 than by writers who are skilled L1 writers.

However, for Self-instructions (r = - .16; p = .17) and Structuring (r = -.11; p = .32), the correlations were not significant.

Relations between L1 use and text quality

To determine whether L1 use while writing in L2 is related to text quality, we correlated the proportion of L1 use for each activity with writers' L2 text quality scores. Results indicate that Generating (r = -.37; p < .001), Goal setting (r = -.24; p < .05), Metacomments (r = -.29; p < .01), and Self-instructions (r = -.29; p < .01) are all negatively related to text quality, while Structuring is not (r = -.17; p = .13). This means that for most of the activities included in the analysis, L1 use appears to have a negative effect on text quality. But the question remains whether this is mediated by L2 proficiency and/or by L1 writing skill. Therefore, we determined the effect of L2 proficiency and L1 writing skill on L1 use and text quality by means of a regression analysis. The results of this analysis are presented in Table 6.

Table 6: The regression weights for the influence of L1 use on L2 text quality (R^2 explained variance) for writers with high and low L2 proficiency and writers with high and low L1 writing skill

	L2 proficiency			*L1 writing skill*		
Cognitive activity in L1	*High (se)*	*Low (se)*	R^2	*High (se)*	*Low (se)*	R^2
Self-instructions	-1.19 (.40)	-.27 (.41)	.12	-.29 (.30)	-1.25 (.32)	.46
Goal setting	-1.05 (.74)	-.88 (.55)	.07	-.74 (.84)	-.43 (.42)	.36
Generating ideas	-3.47 (.97)	-1.10 (.64)	.18	-.18 (.57)	-2.97 (.54)	.53
Structuring	-1.29 (.48)	.06 (.36)	.10	-.02 (.31)	-1.05 (.35)	.41
Metacomments	-1.27 (.41)	-.23 (.40)	.13	-.13 (.29)	-1.40 (.33)	.47

The regression analysis indicated that there is a strong negative relation between L2 proficiency, L1 use and text quality, but only for high L2 proficiency writers. This means that for writers with high L2 proficiency, L1 use during L2 writing appears to be detrimental to text quality, while for lower proficiency writers, text quality does not seem to be affected by the language used. This appears to hold for all activities, except for Goal setting, for which no influence was found on L1 use for either strong or weaker L2 proficiency writers. Furthermore, the percentage of variance explained by L2 proficiency is rather small for all activities (R^2 range: .07 to .18, see Table 6).

There also appears to be a strong negative relation between L1 writing skill, L1 use and text quality, but only for less skilled writers. In other words, L1 use appears to be an indicator of low text quality for texts produced by weaker writers, but no such relation was found between L1 use and text quality for skilled writers. Again, this appears to hold for all activities, except for Goal setting. Whether L1 is used to carry out Goal setting does not appear to be an indicator of low L2 text quality for skilled or less skilled writers. Table 6 also contains the percentage of variance explained by L1 writing skill (R^2 range: .36 to .53), which is much larger, in general than the percentage of variance explained by L2 proficiency.

Finally, we carried out a regression analysis in which L2 proficiency and L1 writing skill were both included. Results indicated that the effect of L2 proficiency on the relation between L1 use and text quality disappears, when the two factors are combined. Generating is the only activity for which L1 use still appears to have a negative influence on text quality for writers with a high level of L2 proficiency. This suggests that L1 writing skill appears to have a larger influence on the relationship between L1 use and text quality than L2 proficiency does. This is confirmed by the fact that L2 text quality correlated with L1 writing skill ($r = .69$; $p < .001$) and also with L2 proficiency ($r = .33$; $p < .05$), while L1 writing skill and L2 proficiency did not correlate with each other ($r = .07$; $p = .57$). This suggests that both these factors are separate components which each have their own influence on text quality (see Cumming, 1989; see also Jourdenais, 2001).

In sum, results indicate that almost all writers use L1 while writing in L2 to some extent, although this varies between activities. In addition, L1 use appears slightly more likely to occur in the beginning of the writing process than near the end. Results also indicate that L1 use appears rather stable between tasks for Generating, but varies somewhat more between tasks for the other four activities. Furthermore, L1 use does not appear to be related between activities, as we only found a moderate positive correlation between Self-instructions and Metacomments, while the correlations between the other activities were generally much lower, or not significant. Regarding the influence of L1 use on text quality, we found that both L1 writing skill and L2 proficiency are negatively correlated to L1 use for most activities, but not for all of them. Moreover, the correlation between L1 use and text quality is further influenced by L2 proficiency and by L1 writing skill, but not in the same way. L1 use appears to have a negative influence on text quality for high L2 proficiency writers and for less-skilled writers. For low L2 proficiency writers and high skilled writers no such effect was found. Finally, the influence of L1 writing skill on the relationship between L1 use and text quality appears to be much larger than the influence of L2 proficiency, which suggests that whether writers use their L1 or not does not only appear to be due to the level of their L2 vocabulary.

CONCLUSION AND DISCUSSION

The aim of this study was to determine to what extent certain conceptual activities occurred in L1 during writing in L2, whether L1 use was related to text quality and whether this relationship was mediated by L1 writing skill and L2 proficiency. Based on our findings, a number of conclusions can be drawn. First of all, in line with earlier research, this study has shown that all writers use L1 while writing in L2 to some extent (Knutson, 2006; Wang & Wen, 2002). L1 use varies between writers, although it appears to be somewhat stable between tasks for Generating ideas. In addition, some activities, such as Self-instructions and Metacomments are more likely to occur in L1 than others. Second, we predicted that L1 use would be related between activities. This was only confirmed to some extent by our results, as the correlations between most activities were moderate at best, except for between Self-instructions and Metacomments. Thus in the case of Self-instructions and Metacomments we can conclude that if writers use L1 for one of these activities while writing in L2, they are also likely to use L1 for the other activity. But this does not mean that we can conclude that L1 use is generally a strong writer specific characteristic. Therefore, the question remains why writers use their L1 during L2 writing. It seems likely that it is not so much the fact that L1 use occurs which influences the quality of the text produced, but rather that L1 use is an indication that writers are finding it difficult to orchestrate their cognitive activities at a specific moment during the writing process. Thus when writers experience cognitive overload while writing in L2, this might cause them to revert to L1 use, which in turn can result in a decrease in text quality.

Third, students at all proficiency levels use L1 for a number of activities while writing in L2 (Knutson, 2006; Woodall, 2002). Furthermore, as predicted, L2 proficiency and L1 writing skill both appear to have a negative effect on L1 use, for most activities. This means that skilled writers and writers with higher L2 proficiency are less likely to use L1 while writing in L2 than weaker writers. These results confirm findings of earlier research on the effect of L2 proficiency on L1 use (Jones & Tetroe, 1987; Sasaki 2002; 2004; Sasaki & Hirose, 1996; Wang & Wen, 2002; Wolfersberger, 2003; Woodall, 2002). Jones & Tetroe (1987), for example, found that L1 use decreased when L2 vocabulary was provided (Jones & Tetroe, 1987, p. 55). However, Wang (2003) found that language switching was more frequent for high proficiency writers than for low proficiency writers, which is not in accordance with our findings. Our participants wrote in cognate languages (L1 Dutch & L2 English) while Wang's participants wrote in non-cognate languages (L1 Chinese & L2 English). This might explain the differences between our findings, as Woodall (2002) found different patterns of L1 use for high and low proficiency writers of cognate versus non-cognate languages (Woodall, 2002, p. 16).

Regarding the effect of L1 use on text quality, we can conclude that L1 use generally appears to be negatively related to text quality. This confirms the results of earlier research, which also suggested that L1 use has a detrimental effect on text quality (see Cohen & Brooks-Cason, 2001, p. 180; Knutson, 2006; Sasaki 2002; 2004; Sasaki & Hirose, 1996; Wang & Wen, 2002; Wolfersberger, 2003). On the other hand, the only exception to the rule is a study by Lay (1982) who found that more L1 use resulted in better texts. However, as she gave no information on how text quality or L1 use were measured, it is impossible to determine why her results differ from ours.

Finally, we predicted that both L1 writing skill and L2 proficiency would have a mediating influence on the relationship between L1 use and text quality. Results indicated that when writers with higher L2 proficiency use L1 while writing, this is likely to reduce the quality of the resulting text. However, the reverse holds for L1 writing skill. When less skilled L1 writers use L1 while writing they are more likely to produce poorer quality texts. So, L2 proficiency and L1 writing skill appear to have different effects on the relation between L1 use and text quality. In his study, Cumming (1989) concluded that writing expertise and L2 proficiency*: "accounted for large, but distinctly separate, proportions of the variance in"* L2 text quality (see Cumming, 1989, p. 118, see also Jourdenais, 2001). Our findings appear to support Cumming's conclusion that writing skill and L2 proficiency are different factors which both influence L2 text quality, as we found that L2 proficiency and L1 writing skill both correlated with L2 text quality, but not with each other. Therefore, we agree with Cumming, that writing expertise might well be: *"a central cognitive ability - with second-language proficiency adding to it, facilitating it in a new domain, and possibly enhancing it"* (Cumming, 1989, p. 121).

Further research

One of the advantages of this study is that we included a multiple task per writer design. But a possible limitation is the fact that we only included tasks in a single genre, argumentative essays, which makes it impossible to generalize our findings across genres (see also Wang & Wen, 2002; Woodall, 2002). Therefore, future research should not only include multiple texts per writer, but also several different types of texts, in order to determine whether L1 use differs between types of tasks. Second, it would be wise to include several levels of L2 proficiency or writing skill in future research. Our participants were all first year English students at University, which implies that they had all attained a certain level of L2 proficiency, and therefore, our comparison between writers was really a comparison of writers at a similar level. Comparing L1 use of writers at different levels of L2 proficiency (see for example Wang & Wen, 2002) or comparing how writers L2 proficiency and L1 use change over time (Sasaki, 2004) could prove informative. One might expect that the mediating effect of L2 proficiency on the relationship between L1 use and

text quality would be stronger for writers with a much lower level of L2 proficiency. Third, our analysis only focused on L1 use for conceptual activities. Therefore, we could not test Wang and Wen's (2002) idea that activities which are more closely linked to text content occur less frequently in L1 while writing in L2. This could be investigated in subsequent research by analysing and comparing the proportion of L1 use for both conceptual and linguistic activities. Finally, our results do not permit us to infer causal relations between L1 use and text quality, as our analyses are correlational in nature. This means that studies with experimental manipulations must be set up to further confirm our findings. This could prove challenging, however, as research has indicated that writers are likely to use L1 while writing in L2, even when they are included in an 'L2-only' condition (see for example, Cohen & Brooks-Carson, 2001).

Implications for teaching

The picture that has emerged from this study is that L1 use during L2 writing is a complex phenomenon. Although L1 use appears negatively related to text quality, this effect is also influenced by L2 proficiency and L1 writing skill. L1 use only appears negatively related to text quality for high L2 proficiency or less skilled L1 writers. For other writers, no detrimental effect on text quality was found, which means that it does not matter whether they use L1 regularly while writing in L2 or not at all. This implies that L1 use does not necessarily have to be discouraged for all writers or under all circumstances. Finally, our findings suggest that L1 writing skill appears to influence text quality through its influence on L1 use during writing in L2. Therefore, it might be advisable for teachers to work with students to improve their L1 writing skill before attempting to improve the quality of their L2 writing.

APPENDIX A: An example of an assignment

Downloading Music for Free

The NSU, the National Student Union, is organising a national essay contest, especially for students. You're also taking part. You absolutely want to win. The winning essay will be printed in all the university newspapers, including the U-blad. The U-blad is read by students and employees of the university.

The subject of the essay has already been decided and was described in the U-blad as follows:
Do you pay for the songs that you like? Do you ever buy a CD? Or do you only download the newest hits from internet for free? The discussion concerning the (il)legal downloading of music is still a hot topic and students make up a large part of the population of internet users who download music. That's why the NSU wants to pay attention to this subject in a special edition of the U-blad. We want to hear from students what they think. Decide what you think and send us your response!

Assignment:
Write an essay in which you give your opinion on the question:
"Downloading Music for Free: criminal or should be possible?"

The essay has to meet the following requirements, set by the Jury:
1. Your essay must be (about) half a page in length.
2. You must do your best to convince your readers, readers of university magazines, of your opinion.
3. You must give arguments to support your opinion.
4. Your essay must be structured in a good and logical way.
5. Your essay must look well-cared-for (think of language use and spelling).
6. In your essay you must use at least two extracts from the 'References' (see next page). You must include these extracts in your essay in a meaningful way.

You have 30 minutes to complete this assignment.

Good luck!

References

Downloading music from the internet has become popular, because of the option of downloading music for free. […] Searching for and downloading music for free is permitted in the Netherlands. However providing music yourself for others to download is not permitted. Source: Bondonline, Digitale Consument, December 2004.

Downloading music from the internet causes an increase in CD-sales, according to a report by market researcher Jupiter Media Metrix. The results of the study contradict claims made by the recording industry. […] Music sales dropped last year by five percent, according to the music industry, due to illegal downloading. […] According to the researchers, the music industry should cherish the users of download platforms instead of suing them. Most users are real music fans, who spend more money on music than average. Source: www.tiscali.nl, May 4th 2002.

Recent research, conducted by researchers from Harvard University and the University of North Carolina, has shown that network services, such as KaZaA have very little to do with the decline in audio-CD sales. […] According to the researchers the effect is "statistically even zero." KaZaA-president Nikki Hemming has already stated she is pleased with the study. "It shows that the music industry shouldn't combat us, but should cooperate with us." […] The Recording Industry Association of America (RIAA), […] has stated that the results contradict those of other studies. Spokeswoman Amy Weiss: "Several other studies have shown the opposite; therefore we still believe that people who download music, buy less music in stores." Source: Joost Blokzijl, www.ZDnet.be, April 5th 2004.

You do pay a price when downloading for free. […] And downloading for free is a bit fishy. When you don't provide music for others to download, but you do download illegally uploaded music to your own computer for free, you're applying double standards. The Consumers' organization doesn't condemn downloading for free. Source: Bondonline, Digitale Consument, December 2004.

A recent study, conducted by researchers from Harvard University and the University of California, has shown that (illegally) downloading music barely has any effect on the sales of music-CD's. […] In many cases downloading actually appears to have a positive effect on the sale of CD's. The authors of the research report confirm the opinion of many downloaders, namely that most of the people who download certain types of music wouldn't have bought that music if they couldn't have downloaded it. The RIAA states that the decline in music sales from more than 13 billion in 2000 to 11 billion in 2003 is almost entirely due to illegal downloads. According to others however, the decline is a result of the economic recession in which the world has found itself over the past few years, as well as increased competition from other forms of entertainment, such as DVD's. Source: M.Sturm, www.tweakers.net, March 30th 2004.

'Europeans love on-line music – as long as it's free'. 36% of downloaders say they buy less CD's because they download music for free. Only 10% say that they have started buying more. Source: Forrester Research – Europe, Augustus 2004.

CHAPTER 7

CONCLUSION AND DISCUSSION

Abstract

The main aim of this research was to examine the influence of language on process-product relations in L1 and L2 writing. Therefore, a group of twenty students wrote texts in their L1 and L2 under think-aloud conditions. A comparison of their L1 and L2 writing processes indicated that writing in L2 influences the moment at which cognitive activities occur during the writing process, as well as the moment at which activities appear to have a positive influence on text quality. Furthermore, writers appear to vary their behaviour less between tasks in L2 than in L1, which might be due, in part, to the increased cognitive load that writing in L2 entails. This final chapter contains a summary of the results of the five studies presented in earlier chapters. In addition, some unanswered questions and unresolved methodological issues are addressed. Finally, some implications for teaching practice are also provided.

Keywords: L1 writing; L2 writing; Keystroke logging; Process-product relations; Text production; Think-aloud method; Writing processes

The series of studies presented in this thesis were carried out in order to advance our understanding of the writing process by *linking* process and product characteristics to each other. The underlying question was how the way in which writers function during the writing process influences the quality of the texts they produce. The Descriptive interactive process model, shown in Figure 1 illustrates how the factors that play a role in relations between the writing process and its product, the text, are related (Van den Bergh, Rijlaarsdam, Janssen, Braaksma, Van Weijen, & Tillema, in press, p. 5; see also Rijlaarsdam, Braaksma, Kieft, Couzijn, Janssen, Raedts, Van Steendam, Groenendijk, Toorenaar, & Van den Bergh, 2008). The model shows that task and learner variables can have a mediating influence on process-product relations. The main aim of the present study was to compare how writers write in L1 and L2, in order to investigate the influence of a specific task variable, language, on process-product relations, through its influence on the orchestration of the writing process. To answer this question, we first had to determine whether the results of earlier first language (L1) writing studies also hold for second language (L2) writing, then determine what effect L2 writing has on the orchestration of cognitive activities during the writing process, and finally, determine the relation between L2 writing processes and text quality.

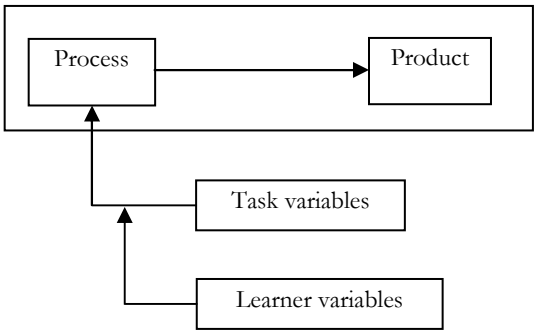

Figure 1: The Descriptive interactive process model (for reading & writing) (Reproduced from: Van den Bergh, et al., in press, p. 5).

Earlier research on L1 writing indicated that the occurrence of cognitive activities varied over time during the writing process, and, more specifically, that the correlation between the occurrence of activities and text quality changed over time as well (see for example, Breetvelt, Van den Bergh, & Rijlaarsdam, 1994; Van den Bergh & Rijlaarsdam, 2001). However, these studies were carried out with high-school pupils instead of students (as was the case in the present study) and they were usually based on a single task per writer, making it impossible to determine the effect of tasks on process-product relations (for a discussion of the methodological implications see Meuffels & Van den Bergh, 2005). Furthermore, a single-task design also makes it impossible to determine whether a writer's approach to the writing task and the quality of the resulting text is better or worse than usual because he's having a good or a bad day. Therefore, in this study, twenty first year university students, who were considered to be rather skilled writers in their L1 Dutch, were asked to write eight short argumentative essays, four in their L1 and four in their L2 English. These participants were trained to work under think-aloud conditions, so that the cognitive activities which they employed while writing could be observed. These data were subsequently examined in several different analyses, in order to determine the effect of writing in L1 and L2 on the orchestration of the writing process and on the quality of the texts produced.

In this concluding chapter, we will summarize the results of the separate studies that were carried out, in relation to the four themes that were presented in Chapter 1. In addition, we will discuss some unanswered questions and unresolved methodological issues, and provide some implications for teaching practice.

Theme 1: Writing in L1 and L2

The main aim of this research project was to determine how the orchestration of the writing process and the quality of the texts produced are influenced by writing in L1 or L2. The fact that we included the same four cognitive activities in our

analyses in Chapters 2 and 3, and also included multiple tasks per writer in each case, enabled us to carry out a comparison of the L1 and L2 results, in order to answer the main question (see Chapter 4).

When we compared the L1 and L2 results, we found significant differences between the temporal distributions of activities between languages, although, in general, the curves were rather similar. The most prominent difference between the two languages, however, was the fact that task effects were much smaller, on average, in L2 than in L1. This means that the orchestration of activities varied far less between tasks in L2 than in L1. In other words, writers' behaviour was generally more consistent in L2 than in L1. Writing in L2 seems, in general, to be an even more complex task than writing in L1 (see for example Jones & Tetroe, 187, p. 53; see also Rijlaarsdam, Braaksma, Couzijn, Janssen, Kieft, Broekkamp, & Van den Bergh, 2005). Thus it seems likely that most of our writers experienced at least a degree of cognitive overload, which appears to have inhibited the variation in the orchestration of cognitive activities over tasks to some extent. Research by Ransdell, Levy and Kellogg (2002), suggests that writers deal with an increase in cognitive load by decreasing their rate of production. This enables them to orchestrate their cognitive activities in the same way as before, and thus prevents a decrease in the quality of the texts they produce (Ransdell, et al., 2002, p. 159). Therefore, it seems likely that the behaviour of the writers in our study became more uniform across tasks, as a means of dealing with the cognitive overload they experienced (Ransdell, et al., 2002; see also Chapter 4). However, even though the orchestration of cognitive activities appeared more stable in L2, the influence of the occurrence of activities on text quality did vary between tasks (see Chapter 4).

In Chapter 6, we examined the effect of language on process-product relations from a different angle, by investigating the effect of L1 use during L2 writing on text quality. In this analysis, we investigated to what extent writers thought aloud in L1 while writing in L2. More specifically, we determined the extent to which different conceptual activities occurred in L1 while writing in L2 and whether this was related to L1 writing skill, L2 proficiency and L2 text quality. Results indicate that L1 use was negatively correlated to both L1 writing skill and L2 proficiency for most activities. In addition, its use is more likely to vary between tasks for some activities than for others. Furthermore, we found that L1 use correlates negatively with L2 text quality, and that this correlation was influenced by L1 writing skill and L2 proficiency. L1 use only appeared to have a negative effect on text quality when it was used by high L2 proficiency writers or by writers with low L1 writing skill. Overall, L1 use appeared to be somewhat more likely to occur in the beginning of the writing process, but we only found evidence of a very gradual decline in L1 use over time. Finally, it seems likely that L1 use does not actually cause a decrease in L2 text quality, but that L1 use is a sign that writers are trying to overcome or prevent cognitive overload while writing in L2. If this

attempt is unsuccessful, then L2 text quality is likely to suffer (see Chapter 6). For a discussion of the implications of using the think-aloud method in this type of research, please see the section on thinking aloud below.

Theme 2: Time as a factor in writing process research

In Chapter 1, we emphasized that the writer's changing task representation is a key element of the writing process, which cannot be ignored in writing process research (see for example Breetvelt, et al., 1994; Rijlaarsdam & Van den Bergh, 2006; Van den Bergh & Rijlaarsdam, 2001; see also Roca de Larios, Murphy, & Marín, 2002). Task representations can change over time during the writing process, but also due to differences between tasks. As a result, we also included an operationalization of the changing task representation in our study, though we did so in two different ways. In Chapters 2, 3, 4, and 6, we reported on research in which the changing task representation was defined according to the moment at which cognitive activities occur during the writing process, in line with earlier research (Breetvelt, et al., 1994, Van den Bergh & Rijlaarsdam, 2001). In all those chapters, we chose to include the segments of think-aloud protocols as a proxy variable for the changing task representations during the writing process. We deemed it appropriate to use segment number instead of time elapsed since the start of the writing process for two reasons. First of all we chose to use segments, because we found a high overall correlation between the number of protocol segments and total writing time (r = .87). In other words, the more time writers needed to write their texts, the larger the number of protocol segments would be (see also Braaksma, Rijlaarsdam, Van den Bergh, & Van Hout-Wolters, 2004; Van den Bergh, et al., in press). Second, and more importantly, we believe that segment number is a more fitting operationalization of the changing task representation than time elapsed since the start of the writing process, because it is more directly related to the occurrence and orchestration of the cognitive activities that occur during writing. Our findings confirmed that the moment at which activities occurred during the writing process was an essential element in our analyses of process-product relations (see Chapters 2, 3, & 4). In those chapters, we analysed the occurrence of individual cognitive activities, in a univariate approach, related to questions raised by Rijlaarsdam and Van den Bergh (1996). However, it seems highly likely that these activities do not occur in isolation, but that their occurrence is connected in some way (see for example, Breetvelt, Van den Bergh, & Rijlaarsdam, 1996; Van den Bergh & Rijlaarsdam, 1999, 2007; Van der Hoeven, 1999). Therefore, in our view, an important and logical next step in writing process research would be to take a multivariate approach (see Rijlaarsdam & Van den Bergh, 1996, p. 112 - 115), in an attempt to determine how these activities are related and how they influence each other during the writing process. In Chapter 5, we took a first step in this direction, by analyzing the moment at which activities occurred within composing episodes.

This was our second operationalization of the changing task representation, which was somewhat different than the operationalization used in earlier chapters. But it was needed, in order to study the influence activities might have on each other. The results of this analysis also confirmed that the moment at which activities occur is crucial, as Planning only appeared to have a positive effect on text quality when it occurred at specific moments during the writing process, namely at the beginning of composing episodes (see Chapter 5).

Theme 3: Variation between tasks

Previous research indicated that both the way in which writers approach the writing process and the quality of the resulting texts can vary strongly between tasks, which means that it is difficult to generalize findings across tasks based on a single-task design (see Meuffels & Van den Bergh, 2005; see also Roca de Larios, Manchón, & Murphy, 2006; Roca de Larios, Marín, & Murphy, 2001; Van den Bergh & Rijlaarsdam, 2001). Therefore, in Chapter 2, we presented results of a study on L1 writing which included four tasks per writer. Because we included multiple tasks per writer, we were able to determine whether the orchestration of cognitive activities varies between tasks in L1 writing. The analysis focused on the occurrence of four cognitive activities: Reading the assignment, Planning, Generating and Formulating. First of all, results indicate that the probability of occurrence of each activity varied over time and between tasks. In addition, the extent to which the occurrence of activities varied between tasks was also time dependent. This means that the extent to which writers vary their behaviour fluctuates over the writing process. Second, the relations between each activity and text quality also appear to vary over time and between tasks, which is in line with findings by Rijlaarsdam and Van den Bergh (1996). In general, tasks appear to influence the strength of the correlation, but not its general pattern. Results for Planning, however, indicate that tasks do influence the correlation, as the correlation between Planning and text quality can be negative or positive depending on the task at hand. Overall, we concluded that variations in the way activities are executed between tasks appear related to variations in text quality, at least for the activities reported in Chapter 2.

In a subsequent analysis, presented in Chapter 3, we determined how time and task affected the orchestration of the writing process in L2 writing. The analysis focused on the same four cognitive activities as in Chapter 2. The effect of time on the occurrence of cognitive activities in L2 writing was rather similar, in general, to its effect on L1 writing. We found that the probability of occurrence of each of the four activities varied over time and between tasks. Furthermore, each activity had a unique correlation with text quality, which varied over time as well. Finally, we concluded that writers appear to have a preference for specific orchestrations of cognitive activities during L2 writing, as their behaviour appeared rather stable over tasks (see Chapter 3).

Theme 4: Interaction between cognitive activities
In Chapter 1, we suggested that cognitive activities do not occur in isolation during the writing process (see for example, Hayes & Gradwohl Nash, 1996, p. 54; Kellogg, 2001, p. 223; Rijlaarsdam & Van den Bergh, 2006; Stevenson, 2005, p. 185; Van den Bergh & Rijlaarsdam, 2007; Van Weijen, 2006). Therefore, in line with questions raised by Rijlaarsdam and Van den Bergh (1996), we sought a means to investigate how activities interact, and enlarging the unit of analysis seemed to have potential in this respect. We decided to focus on conceptual activities such as Planning, because such activities appear able to influence or direct the writing process in some way. Subsequently, an analysis was carried out using composing episodes as the unit of analysis (see also Flower & Hayes, 1981b; Van Weijen, 2006). The analysis revealed that the writing process could be divided into composing episodes, and that the L1 writing process usually contained fewer episodes than were needed when writing in L2. This suggests that writers' attention span appears to have decreased to some extent while writing in L2, which resulted in the need for more episodes than in L1. Furthermore, the writing process appeared to be somewhat more constant in L2, as there were fewer differences between episodes, indicated by the fact that the variance in standard deviations was lower than in L1. Our results also indicated that Planning was the only conceptual activity which had a positive correlation with text quality when it occurred at the start of composing episodes. This suggests, first of all, that Planning is able to influence the quality of the text produced through its influence on the orchestration of the writing process within episodes. Second, this also suggests that writing appears to be a far more hierarchically structured process than was previously assumed, both in research and educational contexts.

METHODOLOGICAL ISSUES

Every study has its set of unresolved issues or methodological problems which might influence the data, affect the interpretation of the results, and/or limit their generalizability, and this study forms no exception. Therefore, some of the unresolved and potentially problematic issues directly related to the present study are discussed below.

Planning
Planning is a notoriously complex activity. Initially, we defined Planning as a combination of self-instructions, goal setting, and structuring activities (see Chapters 2, 3, & 4). After carrying out our L2 analyses, however, we realized that self-instructions might not function in the same way as goal setting and structuring do. Self-instructions (Breetvelt et al., 1994; Rijlaarsdam, 1986) can be seen as a

reflection of the monitor component of the writing process (Flower & Hayes, 1980a; Hayes & Flower, 1980). In addition, the presence of self-instructions can be seen as an indication of a writer's metacognitive control (see for example Tillema, Van den Bergh, Rijlaarsdam, & Sanders, submitted). But the relation between the occurrence of self-instructions and text quality are somewhat unclear. On the one hand, metacognitive control or self-regulation can be a positive characteristic, because, if it is well developed, it enables writers to actively control or steer their writing process in an effective manner (see for example Cumming, 1989; McCutchen, 1988; Graham & Harris, 2000). On the other hand, it might act as a coping mechanism, which does not necessarily have to have a similarly positive effect on the writing process or on text quality. Consequently, if writers regularly use self-instructions during writing, this might be an indication that they are attempting to steer themselves out of trouble (see Chapters 3 & 4). And whether that attempt is successful, probably depends amongst other things, on their level of metacognitive skill, as well as on the effect of other mediating factors such as L1 writing skill and in L2 writing tasks on L2 proficiency. Thus self-instructions can have either a positive or a negative influence on the writing process, whereas the influence of goal setting and structuring is normally positive. Furthermore, goal setting and structuring occurred far less often in our data than self-instructions did (see Chapter 4), which means that the positive influence of these two activities on the writing process would likely be overshadowed by self-instructions if they were all included in one general Planning category. That is why we analyzed the effect of self-instructions and the other types of planning separately in our analyses in Chapters 5 and 6. The results of the analysis of composing episodes in Chapter 5, confirmed that self-instructions appear to play a different role during the writing process than was previously assumed. Self-instructions correlated negatively with text quality when they occurred at episode boundaries, while the correlation between planning (goal setting & structuring) and text quality at episode boundaries was positive (see Chapter 5). Therefore, including composing episodes as the unit of analysis seems to be a promising tool to further our understanding of the way in which cognitive activities influence each other during the writing process.

Thinking aloud

Over the years, research has often indicated that using the think-aloud method in writing research is not without drawbacks (see for example, Ericsson, 1998; Janssen, Van Waes, & Van den Bergh, 1996; Manchón, Murphy, & Roca de Larios, 2005; Roca de Larios, et al., 2006, Smagorinsky, 1994). However, to date, the potential drawbacks of using the method do not appear to outweigh its benefits (e.g., Ericsson & Simon, 1980; Hayes & Gradwohl Nash, 1996; Levy, Marek, & Lea, 1996; Manchón, et al, 2005; Smagorinsky, 1994). Indeed, we agree that: *"The power of protocol analysis lies in the richness of its data. Even though protocols are typically*

incomplete, they provide us with much more information about processes by which tasks are performed than does simply examining the outcome of the process" (Hayes & Flower, 1980, p. 10). Furthermore, investigating the occurrence of conceptual cognitive activities, such as planning, metacomments and generating ideas is almost impossible without resorting to thinking aloud, either concurrently or retrospectively. The same holds when one wishes to study L1 use during L2 writing. There are studies which have collected data on L1 use through self-report questionnaires (e.g., Cohen & Brooks-Carson, 2001; Kobayashi & Rinnert, 1992), but participants were usually unable to report exactly which activities occurred in L1 while writing in L2. As this was the main focus of Chapter 6, including the think-aloud method in this study seemed essential. In addition, concurrent verbalizations are preferable to retrospective interviews, because there is no time-lag between the occurrence of activities and their verbalization (Roca de Larios, et al*.,* 2006, p. 104; see also Ericsson, 1998; Hayes & Gradwohl Nash, 1996, p. 45).

Potentially, thinking aloud in L2 might seem less reliable or more problematic because of the greater cognitive complexity of L2 writing (Jones & Tetroe, 1987, p. 53). Even so, we agree with Krapels (1990) that protocol analysis is a valuable research tool for L2 writing research (Krapels, 1990, p. 51, see also Jourdenais, 2001; Manchón et al., 2005). Moreover, participants in our study (Chapter 6) were free to verbalize their thoughts in L1 or L2 while writing in both languages.

Table 1 contains some examples of situations in which writers in our study switched to their L1 while thinking-aloud in their L2. These examples show that L1 use occurred regularly in short bursts or comments while thinking aloud in L2. Moreover, this does not appear to greatly disrupt writers' trains of thought (see Table 1). The fact that writers frequently resorted to L1 while thinking-aloud in L2 instead of falling silent, indicates in our view that they did not appear to have much difficulty verbalizing their thoughts while writing in L2. But perhaps a more qualitative comparison of writers' thinking aloud in L1 and L2 might bring certain differences to light. Finally, research comparing L2 writing in think-aloud and silent conditions with the use of keystroke logging or alternatively including dual- or triple-task-designs (see for example Kellogg, 1988, 1994; Poilat 1999; Ransdell, et al., 2002) might be able to put an end to this debate.

Adaptivity

The results of our L1 research indicate that the extent to which writers plan during the writing process varies both within and between tasks, and furthermore, that the moment at which planning occurs correlated positively with text quality (see Chapter 2). However, we cannot claim, based on these results, that writers *intentionally* varied their planning behaviour during or between tasks. In other words, whether writers adapt to changing task demands by actively increasing or

decreasing their allocation of available resources to certain activities, such as planning, remains unclear. On the other hand, it does seem very likely that writers adapt their writing behaviour to suit the demands of different writing tasks (see for example, Kellogg, 2001, p. 221; Ransdell, et al., 2002; Roca de Larios, Manchón, Murphy, & Marín, 2008, p. 31; Rijlaarsdam & Van den Bergh, 2006; Severinson Eklundh, 1994, p. 215; Van Weijen, Van den Bergh, Rijlaarsdam, & Sanders, 2005b).

Table 1: Examples of switching between languages while thinking-aloud

Protocol segment (participant, task)	*Translation of L1use*
When America (..) in the volgens mij in de nineteen fifties at least sometime around here I'm not sure, (...) when America prohibited the use and sale of alcohol. (1, H)	I think in the
[Reading the assignment:] Surveillance cameras help prevent crime, but they also increase the chance of tracking down the (..) cul (.) prits (..) ik zou niet weten wat dat zijn (.) but they also increase the chance of tracking down the culprits. The cameras are placed in such a way that perpretratots, perpetrators of crimes within the inner city area are almost almost always registered. (4, D)	I haven't a clue what that is
Mobile phones ho! niet met 'n hoofdletter mobile phone use is important in (.) uhm is important but nost not essential to life and (.) to life. (8, F)	Stop! not with a capital letter
However, (..) uhm, there are way too (..) there are (..) veel te weinig way too less? (.) there are way too less (..) there are way too less uhmmm (..) donors (..) donor (..) registrations. (17, C)	really aren't enough
So it is important that (.) you master (.) ik weet niet of dit goed is maar (.) klinkt wel goed this language. (20, A)	I don't know whether this is right but (.) it sounds fine

(.) = 1 second pause; (..) = 2 second pause; Underlined = L1 (Dutch)

This is further supported by the differences in between-task variation that we found in our comparison of L1 and L2 writing in Chapter 4, and by the changes in process-product relations that we found in L2, despite writers rather stable behaviour between tasks. It seems logical that adapting to changing task demands is only possible if sufficient attentional resources are available. If such resources are limited for whatever reason, then a writer's ability to adapt to the task at hand is likely to suffer (Alamargot & Chanquoy, 2001, p. 176). The writers in the present study varied their behaviour strongly between tasks while writing in L1, but did not do so to the same extent when writing in L2. First of all, this suggests that writers found writing in L2 more complex than writing in L1. This is in line with the idea that writing in L2 is cognitively more demanding than L1 writing due to the increase in task complexity that writing in L2 entails (e.g., Jones & Tetroe, 1987). Second, it suggests that writers experienced a reduction in available attentional resources while writing in L2, particularly for conceptual processes, such as planning and generating ideas. Therefore, writers probably had more working memory resources available for flexible process execution in L1 than in L2. Alternatively, writers' behaviour might have become more uniform because they were enrolled in a course on academic writing in English at the time of the experiment, while no such instruction was provided at the same time in L1. Therefore, experimental research is needed to confirm this cognitive overload hypothesis, and to determine whether this increased flexibility in process execution is indeed a sign of intentional adaptive behaviour (see also Chapter 5).

Keystroke logging

As mentioned in the previous section, the think-aloud method, although very useful, is not perfect. That is why we also collected keystroke logging data in this study. Over the past 15 years or so, the use of keystroke logging software has increased enormously in writing process research. Programs such as Inputlog (Leijten, 2007; Leijten & Van Waes, 2006) and others (see Spelman Miller, 2005; Spelman Miller & Sullivan, 2006) have made it possible to analyse on-line writing processes in far more detail than was previously possible (see Sullivan & Lindgren, 2006, for an overview of recent research). In our study, keystroke logging data was collected for all writing tasks as well, using Inputlog (Leijten & Van Waes, 2006), for two reasons. First of all, we used the logging data as a back-up tool, to complement the think-aloud protocols, which proved to be a very successful combination. The combination of thinking aloud and detailed keystroke logs enabled us to produce richer and more accurate think-aloud protocols than was previously possible based on audio- or video-recordings alone. Second, we hope to analyse the keystroke logging data themselves as well, in future, in order to determine whether process-product relations can be examined through keystroke

log data analysis in a similar way as was done by means of protocol analysis in this study.

In previous research, pauses that occur during the writing process, have often been interpreted as a sign that cognitive processing of some kind, such as planning, is taking place (see for example, Levy & Ransdell, 1995, 1996; Schilperoord, 1996; Ransdell, et al., 2002; Van Waes & Schellens, 2003). Although we agree that pausing is very likely to be related to cognitive processing, it is impossible to determine exactly what is happening with any real accuracy when a writer pauses, based on keystroke logging data alone (see also Spelman Miller, 2005, p. 309), as pauses in keystroke logging data are to a great extent open to interpretation. In the past 15 years or so, research has attempted to solve this problem by using the moment at which pauses appear during writing to aid their interpretation. For example, if writers pause at a boundary between two sentences or between two paragraphs, then it is likely that they are either evaluating the textual unit they have just completed, or that they are planning how to add the next sentence or paragraph onto the text-produced-so-far (see for example, Schilperoord, 1996, p. 12). But whether this is really the case, can only be determined by combining keystroke logging data with other research tools such as thinking-aloud. The same holds for the interpretation of the distribution of pauses over time. Perhaps pauses at the start of the writing process are more likely to be related to planning, while pauses near the end of the writing process might reflect evaluating or rereading the text-produced-so-far. It seems important, therefore, to study the link between pauses in keystroke logging data and the occurrence of cognitive activities during the writing process. It seems probable that planning and generating are more likely to occur during pauses between paragraphs than in pauses between words. But whether that is the case has yet to be investigated. Clarifying what actually happens during such pauses, will clear the way for researchers to begin to establish a meaningful relationship between keystroke logging data on the one hand, and product characteristics, such as discourse structure and text quality, on the other. As a result, we remain convinced that keystroke logging in general, and Inputlog in particular, is an excellent non-obtrusive tool for recording and studying on-line writing processes (see also Spelman Miller, 2005, p. 311), although a combination of keystroke logging with concurrent thinking aloud appears to be the most fruitful approach for studying process-product relations.

Text quality assessment

A key methodological issue in studies which focus on process-product relations is the assessment of the quality of the products or texts produced. Text quality assessment is notoriously difficult, because it is a process which is inherently subjective to a certain extent. As a result, text quality assessment is influenced

strongly by the raters or judges who carry out the assessment (see for example, Coffman, 1966; Schoonen, 1991; 2005; Van den Bergh, De Glopper, & Schoonen, 1988; Wesdorp, 1974). In addition, it is complicated even further by the fact that text quality is known to vary greatly within writers (between tasks) (see Coffman, 1966; Schoonen, 1991; 2005; Van den Bergh, et al., 1988), which makes it both impossible and unadvisable to determine a writer's skill accurately based on the output of a single writing task. Finally, research has indicated that it is generally not possible to compare text quality scores in several languages (see Stevenson, 2005; Stevenson, Schoonen, & De Glopper, 2006). This appears to be due to the fact that it is extremely hard to assess text quality in a similar manner across languages (see Roca de Larios, et al., 2002, for an overview), for example because of the simple fact that writers are far more likely to produce spelling errors in L2 than in L1, which is likely to have a negative effect on L2 text quality scores.

The way in which we defined text quality in this study is inherently Dutch in nature, as it was partly based on assessment methods from earlier studies by other Dutch researchers (see for example, Blok, 1986; Breetvelt, et al., 1994; Rijlaarsdam, 1986). Consequently, a valid question would be whether our definition of text quality is comparable to the English definition of text quality. The answer is, probably not. Research by Burrough-Boenisch (2002), for example, suggests that sentences in Dutch texts are often shorter, on average, than those in English texts (Burrough-Boenisch, 2002, p. 61; see also Burrough-Boenisch, 2004, p. 133 - 137). Other text characteristics that differ between the two languages, include their definitions and use of paragraphs (see Burrough-Boenisch, 2002, p. 63), and the fact that hedging is a far more dominant feature of English texts than Dutch ones (Burrough-Boenisch, 2002, p. 76). Hence, it seems reasonable to conclude that Dutch and English have different conventions for what makes texts acceptable or high in quality. However, we wanted to compare text quality scores across languages, which meant that we had to choose a single set of criteria that can be applied to texts in both languages. In Chapter 4 we reported on an attempt to resolve this L1 – L2 comparison issue, by developing an analytical rating scheme which focused more on the structure, argumentation and content of the text produced than on spelling or other surface level linguistic features. The idea was that structure and argumentation were likely to be stable feature in texts written by the same person in L1 and L2 (see Sanders & Schilperoord, 2006). Furthermore, we translated eight average L1 texts (the benchmark essays used in the L1 holistic assessment) into L2 and included those in the L2 text quality assessment. But translating the L1 essays into L2 also proved difficult, as the errors that occurred had to be translated adequately as well. The reason for translating the essays was that the translated L1 essays were expected to be rated more highly than the L2 benchmark essays. If this turned out to be the case, then it would enable us to determine how we could transform the L1 and L2 scores onto the same scale, thus

making them more or less comparable between languages. Results indicated, however, that two of the translated L1 benchmark essays were assigned much lower scores in L2 than we had expected. Therefore, there was no consistent difference between the L1 and L2 essay scores across tasks, making it impossible to place them on the same scale and compare them directly. It also meant that a comparison of text quality scores in both languages within writers was also not possible.

Although this result prevented us from comparing product quality directly across languages, a comparison of process-product relations in L1 and L2 was still possible, as this was based on the correlations between the occurrence of cognitive activities and text quality in both languages. It was not possible to determine to what extent the L1 essays were better than the L2 ones, because we could not calculate a common average score on the same scale. But a comparison across languages was still possible. However, the search for a means of comparing the quality of texts produced by the same writer in several languages continues. Potential solutions include asking bilingual teachers to judge both the L1 and L2 essays, translating all the essays into L1 or L2 before assessing their quality, or focusing assessment even more on the content of the texts in both languages rather than on their linguistic characteristics. Finally, we would like to stress that the way in which we defined text quality in this study was nevertheless a rather successful one, given the outcome of our research. The results of our analysis in Chapter 6, for example, indicated that the influence of L1 writing skill, defined as a writer's average L1 text quality score, had more influence on L2 text quality than L2 proficiency did.

SUGGESTIONS FOR FUTURE RESEARCH

We have several suggestions for future research, based on the findings discussed in Chapters 2 to 6. Naturally, these are also related to some extent to the methodological issues which were raised in the previous section.

Task-related issues

The most important implication of our research in methodological terms is the fact that writers can vary their behaviour to such a great extent between tasks (see Chapters 2 & 4), that research based on a single task per writer per condition should be avoided at all cost. Otherwise it will be impossible to separate the effect of task-variables, such as the topic of the task, the genre (e.g., narrative versus argumentative essays) the language in which the task is carried out (e.g., L1 versus L2), or the medium (e.g., pen & paper versus computer) from differences between

the experimental conditions. This holds both for studies which measure on-line processing as well as for studies which measure the quality of the texts produced.

In line with this, it is important to mention that we were unable to adhere to this rule in our analysis in Chapter 5, due to time constraints. Results of our analysis indicate that planning at episode boundaries has a positive effect on text quality (see Chapter 5). However, the research design in this study consisted of a single task per condition, which means that the results cannot be generalized beyond the scope of the original study. Therefore, future research will be carried out to investigate whether the results of Chapter 5 on composing episodes can be generalised over tasks. Second, the results of all the analyses in this project only hold for argumentative essays, not for other genres such as narratives, descriptive texts or letters and e-mails. This holds particularly for the study on L1 use during writing in L2 (Chapter 6), as results of earlier research indicated that patterns of L1 use differed for argumentative and narrative texts (Wang & Wen, 2002; Woodall, 2002). Thus a replication of the present study with narrative or descriptive tasks might help determine whether our results also hold for other writing contexts.

Changing task-representations

As mentioned earlier, the results of our analyses in Chapters 2, 3, and 4 clearly confirmed those of earlier research, which indicated that: *"If the moment an activity is employed is left out of the analysis, hardly any relation can be found between cognitive activities on one hand and text quality on the other. Therefore, the moment a cognitive activity takes place is crucial."* (Rijlaarsdam & Van den Bergh, 2006, p. 42, see also Breetvelt, et al., 1994, 1996; Rijlaarsdam & Van den Bergh, 1996; Roca de Larios, et al., 2002; Van den Bergh & Rijlaarsdam, 1999, 2001, 2007; Van den Bergh, et al., in press). Therefore, it is unthinkable, in our opinion that writing process research in the future will be carried out without including production time, protocol segments or some other proxy variable for the changing task representations that writers construct during the writing process. In addition, modelling the occurrence of cognitive activities over time can best be done, in our view, by means of multilevel modelling. The added advantage of this method is that it takes into account the hierarchical nature of the data, and thus also allows researchers to identify both variation between writers and tasks, as well as how this variation fluctuates over time (see for example, Quené & Van den Bergh, 2004, 2008; Van den Bergh, et al., in press, p. 33; Van den Bergh & Rijlaarsdam, 1996).

Other cognitive activities

As mentioned before, four cognitive activities were included in the main analyses for this study: Reading the assignment, Planning, Generating ideas and Formulating. These four activities were chosen for a number of reasons. First of all, we included them because were interested in comparing the effect of writing in

L1 versus L2 on the orchestration of conceptual activities, such as Planning and Generating, and linguistic activities, such as Reading and Formulating. Second, we included these four activities because earlier research has indicated that they each have a clear link with text quality, at least for high school writers (see for example, Breetvelt, et al., 1994; Van den Bergh & Rijlaarsdam, 1999, 2001, 2007) and 12-year olds (Van der Hoeven, 1996a, 1999). Third, we chose to include these four activities because they each make a unique contribution to the writing process (see Chapters 2, 3, & 4). Reading the assignment, for example, provides information on the extent to which writers use the assignment during the writing process. Planning was included because of its influential role during the writing process as a whole, while Generating and Formulating were chosen because they form the core components of the writing process.

Results from Chapters 2, 3, and 4 indicate that, in general, the occurrence of each activity varies over time and over tasks. In addition, variation in the orchestration of these activities seems to influence the quality of the texts produced. We do not know, however, whether the results of the analysis for L1 and L2 hold for all of the other cognitive activities included in our model, such as metacomments, reading the text produced so far, revising and evaluating. Although we suspect that these activities are likely to show the same trend as the four activities that we did analyze, further research must be carried out to determine whether this is in fact the case. In addition, in Chapter 6, we only focused on the use of L1 while writing in L2 for conceptual activities. Wang and Wen (2002) found that L1 use was more likely for activities which are closely linked to the context of the text than for conceptual activities. Future research should test this claim by analysing and comparing the proportion of L1 use for both linguistic and conceptual activities.

Comparison with earlier research

There are at least two points which need to be addressed regarding the relation between our research findings and those of earlier studies. The first point is the fact that our L1 results differed from those of an earlier study by Levy & Ransdell (1995, 1996). Our L1 results (Chapter 2) confirmed the results of several earlier studies (Breetvelt et al., 1994; Van den Bergh & Rijlaarsdam, 2001), even though those studies were carried out with younger participants (high-school pupils). However, our results contradicted findings by Levy & Ransdell, who concluded that writers' behaviour was so stable between tasks that they coined the term *"writing signatures"* (Levy & Ransdell, 1995, 1996). Our results, on the other hand, indicate that writers' behaviour is far more likely to vary between tasks, at least in L1. There are a number of methodological differences between our study and theirs, which might explain the differences in these results. First of all, writers' stable behaviour between tasks in Levy and Ransdell's study might be due to the

fact that they trained their participants to work under think-aloud conditions in a laboratory setting for four weeks before collecting data for their analyses. Training participants to work under think-aloud conditions for such a long period of time might have some benefits from a methodological standpoint, but it also increases the risk that students will see the writing task as a 'trick' they must perform, which might cause a decrease in motivation. Therefore, the within-writer stability that Levy and Ransdell found might be a sign that students developed a coping strategy for dealing with the writing tasks as efficiently as possible, instead of a sign that writers also have a stable writing signature outside of the laboratory setting.

Second, Levy and Ransdell only coded for the occurrence of Planning, Text generating, Revising and Reviewing, while we coded for eleven different cognitive activities (see Chapter 2). If we had aggregated our data into the same four categories, then we might have found more stable patterns within writers as well, as it seems logical that reducing the number of categories in the analysis increases the chance of finding similarities or stable patterns. The same holds for the different definitions of Planning that we employed. Levy and Ransdell included pauses in their planning category, while our definition was far more content based. Their definition is in line with other studies which suggested that a pause indicates that higher level processing, such as planning, is taking place (see for example, Schilperoord, 1996; Ransdell, et al., 2002, p. 144; Van Waes & Schellens, 2003). Furthermore, Schilperoord (1996) found that writers consistently paused longer at boundaries between paragraphs than between sentences or words, which was a rather stable pattern (Schilperoord, 1996, p. 94 – 95). Hence, it seems likely that we would have found more stable patterns as well if we had included pauses as a sign of planning in our definition. Whether this is the case will be investigated in future research.

Third, writers in Levy and Ransdell's study were trained to think aloud and react to a tone while writing. Even though they report that this combination of tasks did not have a negative effect on the way writers' wrote, it could well be that their writers experienced a higher cognitive load while writing than our writers did, and thus had less room available to vary their writing behaviour between tasks. This is in line with our results which indicate that writers appear to vary their behaviour less between tasks while writing in L2. Furthermore, Levy and Ransdell's writers were given very unrestricted tasks, in the sense that they were free to decide on the genre (e.g., narrative, essay, or letter) of the text. Apart from the fact that giving writers free-reign to decide the genre of their text makes a comparison between texts rather difficult, it might also make the task more cognitively demanding. Overall, the combination of an unrestricted task, thinking aloud and a reaction time task might have caused writers to experience an increase in cognitive load, comparable to what our writers experienced when writing in L2, thus reducing their ability or their desire to vary their behaviour between tasks. Clearly,

further research must be carried out to determine whether the differences between these two studies are due to methodological or other factors. However, the between task variation that we came across seems, intuitively, more plausible, as it seems natural that writers will (attempt to) adapt their behaviour to different writing tasks. Therefore, we predict that if we were to replicate Levy and Ransdell's (1995) study, or if we were to reanalyse our own data using their coding scheme and method of analysis, we would also find substantial between-task variation, rather than very stable writer-specific patterns, although the extent to which writers vary their behaviour across tasks is likely to be a writer characteristic (see also Tillema, et al., submitted).

The second point that deserves to be discussed, concerns the comparison between L1 and L2 writing, presented in Chapter 4. Jones & Tetroe (1987) suggested that studies based on quantitative comparisons usually find differences between L1 and L2 writing, while those which focus on qualitative comparisons usually conclude that the two are rather similar (Jones & Tetroe, 1987, p. 39, see also Cumming, 1989, p. 83; Roca de Larios et al., 2002; Silva, 1993). Our findings appear to confirm this difference between quantitative and qualitative studies. On the one hand, our quantitative analysis revealed significant differences between the temporal distributions of cognitive activities over the writing process in L1 and L2. But on the other hand, a qualitative comparison of these distributions revealed that they were actually rather similar (see Chapter 4). The main quantitative difference we found was a reduction in the variation of temporal distributions of activities between tasks, which appears to be a result of a decrease in writers' ability to adapt to changing task demands as a result of the increased complexity of writing in L2. Hence, similarities between L1 and L2 might be due to the fact that writing appears to a certain extent to be a *"central cognitive ability"* (see Cumming, 1989, p. 121), while differences are related to the effect writing in L2 has on this ability.

Inferring causal relations

Strictly speaking, the results of the analyses in Chapters 2 through 6 are all correlational in nature. This means, as we have stated repeatedly in earlier chapters, that although causal relations between the occurrence of cognitive activities and text quality seem logical and rather probable (see also Rijlaarsdam & Van den Bergh, 2006, p. 51), they have not actually been demonstrated. Therefore, experimental research must be carried out in order to provide evidence that planning at specific moments during the writing process actually causes an increase or decrease in text quality. Such experiments could focus, for example, on stimulating or preventing writers from using a specific activity such as planning in different phases of the writing process, for example by increasing writers' cognitive load through the use of dual- or triple-task designs (see for example Kellogg, 1988, 1994; Poilat 1999; Ransdell, et al., 2002). By comparing the resulting text quality

scores and other product characteristics such as text length and discourse structure across conditions, the effect of planning at specific moments on text quality might hopefully be confirmed. The findings presented in this thesis based on think-aloud protocols provide the opportunity to conduct on-line text production research in future, using less intrusive methods such as keystroke logging or pause analysis, without the risk of interference which the collection of concurrent verbal protocols can have. Therefore, future research should be set up in which other on-line measures of text production, such as keystroke logging and pause analysis, are compared to and/or combined with the collection of concurrent or retrospective verbal protocols, to further increase our understanding of the complex nature of process-product relations in writing.

Mediating factors

The analyses described in this thesis, focused on determining the effect of the language in which writers produce their texts (L1 versus L2) on process-product relations. The participants in our study were considered to be highly skilled L1 writers, with an intermediate level of L2 proficiency. Our analysis of the effect of L1 use during L2 writing on text quality, indicated that both L1 writing skill and L2 proficiency influenced process-product relations in this case (see Chapter 6). But it is uncertain whether the way in which we operationalized these variables was suitable for the purpose of our study. For example, we measured writers L2 proficiency by means of a vocabulary test and it is debatable whether this was really an adequate representation of L2 proficiency for these learners. It might well be the case that their L2 vocabulary was sufficient, but that other elements of their L2 were less well developed. The same holds for our operationalization of L1 writing skill, for which we used the quality of their L1 texts. Although equating L1 text quality with writing skill is not necessarily wrong, research has suggested that it might be more appropriate to measure various more specific skills, such as planning or revision skill (see for example Van der Hoeven, 1996a). In a study which focused on the writing processes of 12 year olds, Van der Hoeven indicated that process-product relations can be influenced by language proficiency, planning skill and revision skill (see Van der Hoeven, 1996a, 1996b; see also Rijlaarsdam, et al., 2005, p. 132 - 133), as well as by evaluation skill (Van der Hoeven, 1999). Writers with a higher level of revision skill, for example, were more likely to revise at appropriate moments during the writing process than writers with a lower level of revision skill (Van der Hoeven, 1996b, p. 151). The main point here seems to be that cognitive activities only appear to have a positive influence on the quality of the texts produced when they occur at 'the right time' during the writing process. Van der Hoeven's (1996a, 1996b, 1999) findings suggest that this might be related to the development of cognitive skills. Writers are only likely to plan at the right time during the writing process if they have a sufficiently developed level of

planning skill to know when that 'right time' is. The same appears to hold for revision skill as well. Therefore, it seems important to further investigate the influence of cognitive skills on the orchestration of cognitive processes by more advanced writers, such as the ones who participated in the present study. In addition, research by Ransdell, et al. (2002) indicated that working memory capacity also influences writers' process execution (see Ransdell, et al., 2002), although the influence of other factors such as fluency on L2 writing remains somewhat unclear (Stevenson, 2005). Therefore, it seems important to further investigate the influence of such factors on process-product relations in more detail in future.

In sum, we can conclude that process-product relations change in a number of ways as a result of writing in L2. First of all, our findings indicate that writing in L2 influences the moment at which cognitive activities occur during the writing process (see Chapters 3 & 4). Second, we found that the moment at which these activities appear to contribute positively to text quality changes when writing in L2 (see Chapter 4). Finally, writing in L2 appears to result in a decrease in the variation of the orchestration of cognitive activities between tasks compared to writing in L1. In other words, when writing in L2, writers appear to vary their behaviour less between tasks than in L1 (see Chapter 4), perhaps because an increase in cognitive load inhibits them from doing so.

IMPLICATIONS FOR TEACHING

Although the main purpose of this study was not directly related to education or teaching practices, some of our findings might be beneficial for writers and teachers, as well as those who are learning to write. Therefore, we would like to offer some suggestions for writing in educational settings.

First of all, our results indicate that writers' behaviour can fluctuate strongly between tasks, particularly in L1 writing (see Chapters 2, 3, & 4). As a result, text quality can also vary, which means that there is no guarantee that a writer who produces a good essay on one day will also write a good essay a week later. Therefore, the assessment of writing skill should always give writers the opportunity to write multiple texts, in order to minimize the influence of a single really good or really poor quality text on the outcome of the assessment as a whole.

Second, Flower and Hayes proposed that writing is a complex task during which a writer is constantly forced to juggle with constraints (Flower & Hayes, 1980a, p. 40) Over the years, subsequent research has consistently confirmed this idea (see for example Alamargot & Chanquoy, 2001; Kellogg, 1994, 2001; Van den Bergh & Rijlaarsdam, 1999). Our results confirm that this also holds for highly proficient L1 writers (see Chapter 2). Furthermore, our results suggest that L2

writing is a cognitively even more demanding task than L1 writing (see Chapters 3 & 4). This is important to keep in mind during L2 writing instruction, as it means that writers are learning with an added handicap as it were. In light of this, it seems advisable to separate writing from learning to write both in L1 and in L2 writing. This means that writers should be given the opportunity to learn to write in guided steps, not only by writing lots of essays.

Third, our findings indicate that the moment at which cognitive activities are carried out is an essential feature that distinguishes successful writers from weaker writers (see Chapters 2, 3, & 4). Therefore, based on our findings, we recommend that teachers use a process approach to writing instruction. Simply telling students when they should plan, formulate or revise is not likely to radically improve the way they write. It seems far more worthwhile to demonstrate how writers deal with all the constraints they encounter while writing, by letting students observe other writers as they work. This can be done, for example by directing students' attention to the presence of composing episodes in the writing process (see Chapter 5). This is advisable, because it provides concrete ways for students to get a handle on their own cognitive processing. The idea behind this, for example, is that if writers pause after completing a paragraph in order to get their bearings as it where and thus determine what they should do next, then that is likely to have a positive effect on text quality. Research by Braaksma, et al. (2004) has shown that observational learning can be an important tool in this respect, as it offers students the chance to separate writing from learning to write. The results of their study indicate that students who observed model writers at work showed more improvement in their orchestration of cognitive activities than students who had to learn by writing texts themselves (Braaksma, et al., 2004, p. 28; see also Braaksma, 2002). Although Braaksma et al.'s findings apply to high-school pupils, research by Van Steendam, Rijlaarsdam and Sercu (2007; 2008) suggests that more proficient student writers, such as our participants, might also benefit from observational learning, particularly in an L2 context (see also Rijlaarsdam, Braaksma, Couzijn, Janssen, Raedts, Van Steendam, Toorenaar, & Van den Bergh, 2008). However, it is important to note that it is essential to also provide students with examples of less successful writers, because students are likely to learn more from others' mistakes than from observing highly skilled writers.

Fourth, a longstanding idea in L2 learning contexts is the idea that using one's first language while writing in L2 has a negative effect on the quality of the texts produced. Research has indicated that this is not necessarily the case (see for example Jones & Tetroe, 1987; Knutson, 2006, p. 88; Uzawa & Cumming, 1989, p. 187; Wang & Wen, 2002). The results of our study also indicate that using L1 while writing in L2 does not always have a negative influence on text quality (see Chapter 6), although the picture is rather complex. On the one hand, using L1 seems negatively related to L2 text quality for writers who are less skilled writers in their

L1, while no such effect was found for skilled L1 writers. In other words, skilled L1 writers are free to use their L1, while it seems less advisable for less skilled L1 writers. On the other hand, we found that L1 use was related to L2 proficiency as well. L1 use appeared to have a negative effect on L2 text quality for writers with a high level of L2 proficiency, but not for low L2 proficiency writers. Furthermore, our findings also indicate that L1 use during L2 writing is influenced more by L1 writing skill than by L2 proficiency. This suggests that whether writers switch to their L1 while writing in L2 is not merely caused by a lack of L2 proficiency. There are some indications, for example, that writers switch to their L1 while writing in L2 in order to reduce the risk of cognitive overload and thus keep the writing process on track (see for example, Cohen & Brooks-Carson, 2001; Knutson, 2006; Qi, 1998; Woodall, 2002; see also Chapter 6). It might thus be beneficial to give students the opportunity to practice using their L1 during L2 writing tasks. Again, observational learning might be a useful tool in this respect (Braaksma, et al., 2004). By observing how other writers incorporate their L1 in their L2 writing process, students can learn to use their L1 as a tool in their own writing.

Finally, our research suggests that L1 writing skill might have a larger influence on the quality of L2 texts than L2 proficiency (see Chapter 6). This suggests that writing expertise might be a language-independent cognitive ability (Cumming, 1989, p. 121), which might hold particularly for conceptual activities such as planning and generating ideas. If that is indeed the case, then it seems advisable that students and teachers focus their energy first on improving students' L1 writing skill, before students attempt to write in their L2.

REFERENCES

Akyel, A. (1994). First language use in EFL writing: planning in Turkish vs. planning in English. *International Journal of Applied Linguistics, 4*(2), 169-197.

Alamargot, D., & Chanquoy, L. (2001). Through the models of writing. In G. Rijlaarsdam (Series Ed.), D. Alamargot & L. Chanquoy (Vol Eds.), *Studies in Writing Vol. 9,* (Dordrecht: Kluwer Academic Publishers.

Arndt, V. (1987). Six writers in search of texts: a protocol-based study of L1 and L2 writing. *ELT Journal, 41*(4), 257-267.

Beare, S. (2000). *Differences in content generating and planning processes of adult L1 and L2 proficient writers.* University of Ottawa.

Beare, S., & Bourdages, J. S. (2007). Skilled writers' generating strategies in L1 and L2: an exploratory study. In G. Rijlaarsdam (Series Ed.), M. Torrance, L. Van Waes & D. Galbraith (Vol Eds.), *Studies in Writing Vol. 20, Writing and Cognition: Research and Applications* (pp. 151 - 161). Amsterdam: Elsevier.

Bereiter, C., & Scardamalia, M. (1987). *The Psychology of written composition.* Mahwah, NJ: Lawrence Erlbaum Associates.

Blok, H. (1986). Essay rating by the comparison method. *Tijdschrift voor Onderwijsresearch, 11*(4), 169 - 176.

Braaksma, M. A. H. (2002). *Observational learning in argumentative writing.* Amsterdam: Graduate School for Teaching and Learning, University of Amsterdam.

Braaksma, M., Rijlaarsdam, G., Van den Bergh, H., & Van Hout-Wolters, B. H. A. M. (2004). Observational learning and its effects on the orchestration of writing processes. *Cognition and Instruction, 22*(1), 1-36.

Breetvelt, I. (1988). *De relatie tussen schrijfvaardigheid en kenmerken van het schrijfproces; een literatuurstudie* [The relationship between writing ability and characteristics of the writing process; a literature study] (SCO Rapport No. 179). Amsterdam: Stichting Centrum voor Onderwijsonderzoek.

Breetvelt, I., Van den Bergh, H., & Rijlaarsdam, G. (1994). Relations between writing processes and text quality: when and how? *Cognition and Instruction, 12*(2), 103-123.

Breetvelt, I., Van den Bergh, H., & Rijlaarsdam, G. (1996). Rereading and generating and their relation to text quality. An application of multilevel analysis on writing process data. In G. Rijlaarsdam & E. Espéret (Series Eds.), G. Rijlaarsdam, H. Van den Bergh & M. Couzijn (Vol Eds.), *Studies in Writing Vol. 2, Theories, models and methodology in writing research* (pp. 10 - 20). Amsterdam: Amsterdam University Press.

Breuker, J. A., Elshout, J. J., Van Someren, M. W., & Wielinga, B. J. (1986). Hardopdenken en protokolanalyse [Thinking aloud and protocol analysis]. *Tijdschrift voor Onderwijsresearch, 11*(5), 241-254.

Broekkamp, H., & Van den Bergh, H. (1996). Attention strategies in revising a foreign language text. In G. Rijlaarsdam & E. Espéret (Series Eds.), G. Rijlaarsdam, H. Van den Bergh & M. Couzijn (Vol Eds.), *Studies in Writing Vol. 2, Theories, models and methodology in writing research* (pp. 170 - 181). Amsterdam: Amsterdam University Press.

Burrough-Boenisch, J. (2002). *Culture and conventions: writing and reading Dutch scientific English* (Vol. 59). Utrecht: LOT.

Burrough-Boenisch, J. (2004). *Righting English that's gone Dutch*. Voorburg: Kemper Conseil Publishing.

Carey, L., J., & Flower, L. (1989). *Foundations for creativity in the writing process: Rhetorical representations of ill-defined problems* (Tech. Rep. No. 32). Berkeley, CA; and Pittsburgh, PA:. Center for the Study of Writing.

Chenoweth, N. A., & Hayes, J. R. (2001). Fluency in writing, generating text in L1 and L2. *Written Communication, 18*(1), 80-98.

Coffman, W. E. (1966). On the validity of essay tests of achievement. *Journal of Educational Measurement, 3*(2), 151 - 156.

Cohen, A. D., & Brooks-Carson, A. (2001). Research on direct versus translated writing: students' strategies and their results. *The Modern Language Journal, 85*(ii), 169-188.

Cook, V. (1996). *Second language learning and language teaching* (Second edition ed.). London: Arnold.

Couzijn, M., Van den Bergh, H., & Rijlaarsdam, G. (2002). *Writing Processes and text quality: effects of L1/L2*. Paper presented at the SIG Writing '02 Conference, Staffordshire, UK.

Cumming, A. (1989). Writing expertise and second language proficiency. *Language Learning, 39*(1), 81 - 141.

Cumming, A. (2001). Learning to write in a second language: Two decades of research. *International Journal of English Studies, 1*(2), 1 - 23.

Degand, L., Demol, A., Hardermann, P., & Perrez, J. (2006). *And, then, so, … Why do L2 writers love connectives?* Paper presented at the 10th International Conference of the Earli Special Interest Group on Writing, University of Antwerp, Belgium.

Di Gennaro, K. (2006). Second language writing ability: Towards a complete construct definition. *Teachers College, Columbia University Working Papers in TESOL & Applied Linguistics, 6*(2), 1 - 17.

Ellis, R. (2005). Planning and task-based performance; theory and research. In N. Spada & J. Hulstijn (Series Eds.), R. Ellis (Vol Ed.), *Language learning and language teaching Vol. 11, Planning and task performance in a second language* (pp. 3 - 34). Amsterdam/Philadelphia: John Benjamins Publishing Company.

Ellis, R., & Yuan, F. (2004). The effects of planning on fluency, complexity, and accuracy in second language narrative writing. *Studies in Second Language Acquisition, 26*(1), 59 - 84.

Ellis, R., & Yuan, F. (2005). The effects of careful within-task planning on oral and written task performance. In N. Spada & J. Hulstijn (Series Eds.), R. Ellis (Vol Ed.), *Language learning and language teaching Vol. 11, Planning and task performance in a second language* (pp. 167 - 192). Amsterdam/Philadelphia: John Benjamins Publishing Company.

Ericsson, K. A. (1998). Protocol analysis. In W. Bechteld & G. Graham (Eds.), *A Companion to Cognitive Science* (pp. 425 - 432). Oxford: Blackwell.

Ericsson, K. A., & Simon, H. A. (1980). Verbal Reports as Data. *Psychological Review, 87*(3), 215 - 251.

Fienberg, S. E. (1980). *The analysis of cross-classified categorical data* (2nd ed.). Cambridge, Mass./London: The MIT-press.

Flower, L. S., & Hayes, J. R. (1980a). The Dynamics of composing: Making plans and juggling constraints. In L. W. Gregg & E. R. Steinberg (Eds.), *Cognitive processes in writing* (pp. 31 -50). Hillsdale, NJ: Lawrence Erlbaum Associates.

Flower, L. S., & Hayes, J. R. (1980b). The Cognition of discovery: Defining a rhetorical problem. *College Composition and Communication, 31*(1), 21-32.

Flower, L., & Hayes, J. R. (1981a). Plans that guide the composing process. In C. H. Frederiksen & J. F. Dominic (Eds.), *Writing: The nature, development and teaching of written communication. Volume 2: Writing: process, development and communication* (pp. 39 - 58). Hillsdale, NJ: Lawrence Erlbaum Associates.

Flower, L., & Hayes, J. R. (1981b). The Pregnant pause: An inquiry into the nature of planning. *Research in the Teaching of English, 15*(3), 229 - 243.

Flower, L., Hayes, J. R., Carey, L. J., Schriver, K., & Stratman, J. (1986). Detection, Diagnosis and the Strategies of Revision. *College Composition and Communication, 37*(1), 16-55.

Friedlander, A. (1990). Composing in English: effects of a first language on writing in English as a second language. In B. Kroll (Ed.), *Second language writing; Research insights for the classroom* (109 - 125). Cambridge: Cambridge University Press.

Godshalk, F. I., Swineford, F., & Coffmann, W. E. (1966). *The measurement of writing ability*. New York: College Entrance Examination Board.

Goldstein, H. (1995). *Multilevel Statistical Models* (second edition ed.). London: Arnold.

Grabe, W. (2001). Notes toward a theory of second language writing. In T. Silva & P. K. Matsuda (Eds.), *On Second Language Writing* (pp. 39 - 57). Mahwah, NJ: Lawrence Erlbaum Associates.

Graham, S., & Harris, K. R. (2000). The role of self-regulation and transcription skills in writing and writing development. *Educational Psychologist, 35*(1), 3 -12.

Graham, S., & Perin, D. (2007). A meta-analysis of writing instruction for adolescent students. *Journal of Educational Psychology, 99*(3), 445 - 476.

Hayes, J. R. (1996). A new framework for understanding cognition and affect in writing. In C. M. Levy & S. Ransdell (Eds.), *The science of writing: Theories, methods, individual differences & applications* (pp. 1 - 27). Mahwah, NJ: Lawrence Erlbaum Associates.

Hayes, J. R., & Flower, L. S. (1980). Identifying the organization of writing processes. In L. W. Gregg & E. R. Steinberg (Eds.), *Cognitive processes in writing* (pp. 3-30). Hillsdale, NJ: Lawrence Erlbaum Associates.

Hayes, J. R., & Gradwohl Nash, J. (1996). On the nature of planning in writing. In C. M. Levy & S. Ransdell (Eds.), *The science of writing: Theories, methods, individual differences & applications* (pp. 29 - 55). Mahwah, NJ: Lawrence Erlbaum Associates.

Hinkel, E. (2002). *Second language writers' text linguistic and rhetorical features*. Mahwah, NJ: Lawrence Erlbaum Associates.

Janssen, D., Van Waes, L., & Van den Bergh, H. (1996). Effects of thinking aloud on writing processes. In C. M. Levy & S. Ransdell (Eds.), *The science of writing: Theories, methods, individual differences & applications* (pp. 233 - 250). Mahwah, NJ: Lawrence Erlbaum Associates.

Johnson, R. B. (1997). Examining the validity structure of qualitative research. *Education 118*(2), 282 - 292.

Jones, S., & Tetroe, J. (1987). Composing in a second language. In A. Matsuhashi (Ed.), *Writing in real time; Modelling production processes* (pp. 34 - 57). Norwood, NJ: Ablex Publishing Corporation.

Jourdenais, R. (2001). Cognition, instruction and protocol analysis. In P. Robinson (Ed.), *Cognition & Second Language Instruction* (pp. 354 - 375). Cambridge: Cambridge University Press.

Kellogg, R. T. (1988). Attentional Overload and Writing Performance: Effects of Rough Draft and Outline Strategies. *Journal of Experimental Psychology: Learning, Memory and Cognition, 14*(2), 355 - 365.

Kellogg, R. T. (1994). *The Psychology of Writing*. Oxford: Oxford University Press.

Kellogg, R. T. (1996). A model of working memory in writing. In C. M. Levy & S. Ransdell (Eds.), *The science of writing: Theories, methods, individual differences & applications* (pp. 57 - 71). Mahwah, NJ: Lawrence Erlbaum Associates.

Kellogg, R. T. (2001). Commentary on Part II: processing modalities and development of expertise in writing. In G. Rijlaarsdam (Series Ed.), D. Alamargot & L. Chanquoy (Eds.), *Studies in Writing Vol. 9, Through the models of writing* (pp. 219 - 228). Dordrecht: Kluwer Academic Publishers.

Kellogg, R. T., Olive, T., & Piolat, A. (2007). Verbal and visual working memory in written sentence production. In G. Rijlaarsdam (Series Ed.), M. Torrance, L. Van Waes & D. Galbraith (Eds.), *Studies in writing Vol. 20, Writing and cognition: Research and applications* (Amsterdam: Elsevier.

Kennedy, M. L. (1985). The Composing process of college students writing from sources. *Written Communication, 2*(4), 434 - 456.

Knutson, E. M. (2006). Thinking in English, writing in French. *The French Review, 80*(1), 88 - 109.

Kobayashi, H., & Rinnert, C. (1992). Effects of First Language on Second Language Writing: Translation versus Direct Composition. *Language Learning, 42*(2), 183 - 215.

Krapels, A. R. (1990). An overview of second language writing process research. In B. Kroll (Ed.), *Second language writing; Research insights for the classroom* (37 - 56). Cambridge: Cambridge University Press.

Lally, C. G. (2000). First language influences in second language composition: The effect of pre-writing. *Foreign Language Annals, 33*(4), 428 - 432.

Lay, N. D. S. (1982). Composing processes of adult ESL learners: a case study. *TESOL Quarterly, 16*(2), 406.

Leijten, M. (2007). *Writing and speech recognition; observing error correction strategies of professional writers* (Vol. 160). Utrecht: LOT

Leijten, M., & Van Waes, L. (2006). Inputlog: New perspectives on the logging of on-line writing processes in a Windows environment. In G. Rijlaarsdam (Series Ed.), K. Sullivan & E. Lindgren (Vol Eds.), *Studies in Writing Vol. 18, Computer key-stroke logging and writing: methods and applications* (pp. 73-94). Oxford: Elsevier.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.

Levy, C. M., Marek, J. P., & Lea, J. (1996). Concurrent and retrospective protocols in writing research. In G. Rijlaarsdam & E. Espéret (Series Eds.), G. Rijlaarsdam, H. Van den Bergh & M. Couzijn (Vol Eds.), *Studies in Writing Vol. 2, Theories, models and methodology in writing research* (pp. 542 - 556). Amsterdam: Amsterdam University Press.

Levy, C. M., & Ransdell, S. (1995). Is writing as difficult as it seems? *Memory and Cognition, 23*(6), 767 - 779.

Levy, C. M., & Ransdell, S. (1996). Writing signatures. In C. M. Levy & S. Ransdell (Eds.), *The science of writing: Theories, methods, individual differences & applications* (pp. 149 - 161). Mahwah, NJ: Lawrence Erlbaum Associates.

Manchón, R. M., Murphy, L., & Roca de Larios, J. (2005). Using concurrent protocols to explore L2 writing Processes: methodological issues in the collection and analysis of data. In P. K. Matsuda & T. Silva (Eds.), *Second language writing research; perspectives on the process of knowledge construction* (pp. 191 - 205). Mahwah, NJ: Lawrence Erlbaum Associates.

Manchón, R. M., & Roca de Larios, J. (2007). On the temporal nature of planning in L1 and L2 composing. *Language Learning, 57*(4), 549-593.

Manchón, R. M., Roca de Larios, J., & Murphy, L. (2000). An approximation to the study of backtracking in L2 writing. *Learning and Instruction, 10*, 13 - 35.

McCutchen, D. (1988). "Functional Autmaticity" in Children's Writing; A Problem of Metacognitive Control. *Written Communication, 5*(3), 306 - 324.

McCutchen, D. (2000). Knowledge, processing and working memory: Implications for a theory of writing. *Educational Psychologist, 35*(1), 13-23.

McCutchen, D., Covill, A., Hoyne, S. H., & Mildes, K. (1994). Individual differences in writing: Implications of translating fluency. *Journal of Educational Psychology, 86*(2), 256 - 266.

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

O'Brien, T. (2004). Writing in a foreign language: teaching and learning. *Language Teaching, 37*, 1 - 28.

Oostdam, R., & Rijlaarsdam, G. (1995). *Towards strategic language learning*. Amsterdam: Amsterdam University Press.

Pélissier, A., & Piolat, A. (1998). *Are writing processes activated sequentially or in parallel? The appropriate way to use Kellogg's three task method.* Paper presented at the Writing and Learning at the dawn of the 21st Century - Earli SIG- Writing.

Penningroth, S. L., & Rosenberg, S. (1995). Effects of a high information-processing load on the writing process and the story written. *Applied Psycholinguistics, 16*, 189-210.

Perrez, J. (2006). *Connectieven, tekstbegrip en vreemde-taalverwerving: een studie van de impact van causale en contrastieve connectieven op het begrijpen van teksten in het Nederlands als vreemde-taal* [Connectives, text comprehension and foreign language learning: A study of the impact of causal and contrastive connectives on text comprehension in Dutch as a foreign language]. Université catholique de Louvain.

Piolat, A. (1999). Planning and text quality among undergraduate students: Findings and quotations. In G. Rijlaarsdam & E. Esperet (Series Eds.), M. Torrance & D. Galbraith (Eds.), *Studies in Writing Vol. 4, Knowing what to write; Conceptual processes in text production* (pp. 121 - 137). Amsterdam: Amsterdam University Press.

Qi, D. S. (1998). An inquiry into language-switching in second language composing processes. *The Canadian Modern Language Review/La Revue canadienne des languages vivantes, 54*(3), 413 - 435.

Quené, H., & Van den Bergh, H. (2004). On multi-level modeling of data from repeated measures designs: a tutorial. *Speech Communication, 43*, 103 - 121.

Quené, H., & Van den Bergh, H. (2008). Examples of mixed-effects modeling with crossed random effects and with binomial data. *Journal of Memory and Language*, doi:10.1016/j.jml.2008.1002.1002.

Raimes, A. (1994). Language proficiency: writing ability and composing strategies: a study of ESL college student writers. In A. Cumming (Ed.), *Bilingual performance in reading & writing* (pp. 139 - 172). Ann Arbor: John Benjamins Publishing Co.

Ransdell, S., Levy, C. M., & Kellogg, R. T. (2002). The structure of writing processes as revealed by secondary task demands. *L1 - Educational Studies in Language and Literature, 2*(2), 141 - 163.

Rijlaarsdam, G. (1986). *Effecten van leerlingenrespons op aspecten van stelvaardigheid [Effects of peer-feedback on aspects of writing skill]* (SCO-Rapport No. 88). Amsterdam: Stichting Centrum voor Onderwijsonderzoek.

Rijlaarsdam, G., Braaksma, M., Couzijn, M., Janssen, T., Kieft, M., Broekkamp, H., et al. (2005). Psychology and the teaching of writing in 8000 and some words. *Pedagogy - Learning for Teaching; BJEP Monograph series II 3*, 127-153.

Rijlaarsdam, G., Braaksma, M., Couzijn, M., Janssen, T., Raedts, M., Van Steendam, E., Toorenaar, A., & Van den Bergh, H. (2008). Observation of peers in learning to write; Practise and research. *Journal of Writing Research, 1*(1), 53 - 83.

Rijlaarsdam, G., & Van den Bergh, H. (1996). The dynamics of composing - an agenda for research into an interactive compensatory model of writing: many questions, some answers. In C. M. Levy & S. Ransdell (Eds.), *The science of writing: Theories, methods, individual differences & applications* (pp. 107 - 125). Mahwah, NJ: Lawrence Erlbaum Associates.

Rijlaarsdam, G., & Van den Bergh, H. (2006). Writing process theory: a functional dynamic approach. In C. A. MacArthur, S. Graham & J. Fitzgerald (Eds.), *The handbook of writing research* (pp. 41 - 53). NY: Guilford Publications.

Rijlaarsdam, G., in cooperation with Braaksma, M., Kieft, M., Couzijn, M., Janssen, T., Raedts, M., Van Steendam, E., Groenendijk, T., Toorenaar, A., & Van den Bergh, H. (2008). *The yummy yummy case: Learning to write--Observing readers and writers.* Keynote at the 3rd International Writing Research Across Borders Conference, February 22-24, 2008, University of California, Santa Barbara, USA.

Roca de Larios, J., Manchón, R. M., & Murphy, L. (2006). Generating text in native and foreign language writing: a temporal analysis of problem-solving formulation processes. *The Modern Language Journal, 90*(i), 100 - 114.

Roca de Larios, J., Manchón, R. M., Murphy, L., & Marín, J. (2008). The foreign language writer's strategic behaviour in the allocation of time to writing processes. *Journal of Second Language Writing, 17*(1), 30-47.

Roca de Larios, J., Marín, J., & Murphy, L. (2001). A temporal analysis of formulation processes in L1 and L2 writing *Language Learning, 51*(3), 497-538.

Roca de Larios, J., Murphy, L., & Manchón, R. M. (1999). The use of restructuring strategies in EFL writing: a study of Spanish learners of English as a foreign language. *Journal of Second Language Writing, 8*(1), 13-44.

Roca de Larios, J., Murphy, L., & Marín, J. (2002). A critical examination of L2 writing process research. In G. Rijlaarsdam (Series Ed.), S. Ransdell & M. L. Barbier (Vol Eds.), *Studies in Writing Vol. 11, New directions for research in L2 writing* (pp. 11 - 47): Kluwer Academic Publishers.

Sanders, T., Janssen, D., Van der Pool, E., Schilperoord, J., & Van Wijk, C. (1996). Hierarchical text structure in writing products and writing processes. In G. Rijlaarsdam & E. Espéret (Series Eds.), G. Rijlaarsdam, H. Van den Bergh & M. Couzijn (Vol Eds.), *Studies in Writing Vol. 2, Theories, Models & Methodology in Writing Research* (pp. 473 - 492). Amsterdam: Amsterdam University Press.

Sanders, T., & Schilperoord, J. (2006). Text structure as a window on the cognition of writing; How text analysis provides insights in writing products and writing processes. In C. A. MacArthur, S. Graham & J. Fitzgerald (Eds.), *The handbook of writing research* (pp. 386 - 402). NY: Guilford Publications.

Sanders, T., & Van Wijk, C. (1996). Text analysis as a research tool: how hierarchical text structure contributes to the understanding of conceptual processes in writing. In C. M. Levy & S. Ransdell (Eds.), *The science of writing: Theories, methods, individual differences & applications* (pp. 251 - 269). Mahwah, NJ: Lawrence Erlbaum Associates.

Sasaki, M. (2002). Building an empirically-based model of EFL learners' writing processes. In G. Rijlaarsdam (Series Ed.), S. Ransdell & M. Barbier (Eds.), *Studies in Writing Vol. 11, New directions for research in L2 Writing* (pp. 49 - 80): Kluwer Academic Publishers.

Sasaki, M. (2004). A multiple-data analysis of the 3.5-year development of EFL student writers. *Language Learning, 54*(3), 525 - 582.

Sasaki, M., & Hirose, K. (1996). Explanatory variables for EFL students' expository writing. *Language Learning, 46*(1), 137 - 174.

Schilperoord, J. (1996). *It's about time: temporal aspects of cognitive processes in text production*. Amsterdam/Atlanta: Rodopi.

Schoonen, R. (1991). *De evaluatie van schrijfvaardigheidsmetingen; Een empirische studie naar betrouwbaarheid, validiteit en bruikbaarheid van schrijfvaardigheidsmetingen in de achtste groep van het basisonderwijs.* [The evaluation of writing ability measures; An empirical study of the reliability, validity and usability of writing skill measures in sixth grade]Unpublished doctoral dissertation, Universiteit van Amsterdam, Amsterdam.

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

Schoonen, R., Van Gelderen, A., De Glopper, K., Hulstijn, J., Simis, A., Snellings, P., & Stevenson, M. (2003). First language and second language writing: the role of linguistic knowledge, speed of processing and metacognitive knowledge. *Language Learning, 53*(1), 165 - 202.

Severinson Eklundh, K. (1994). Linear and Nonlinear Strategies in Computer-Based Writing. *Computers and Composition, 11*, 203-216.

Shi, L. (2004). Textual borrowing in second language writing. *Written Communication, 21*(2), 171 - 200.

Silva, T. (1993). Toward an understanding of the distinct nature of L2 writing: the ESL research and its implications. *TESOL Quarterly, 27*(4), 657 - 677.

Smagorinsky, P. (1994). Think-aloud protocol analysis, beyond the black box. In C. R. Cooper & L. Brodkey (Series Eds.), P. Smagorinsky (Vol Ed.), *Sage series in Written Communication Vol. 8, Speaking about writing, reflections on research methodology* (pp. 3 - 19). Thousand Oaks, Ca.: Sage Publications.

Snijders, T. A. B., & Bosker, R. J. (1999). *Multilevel analysis; an introduction to basic and advanced multilevel modelling*. London: Sage publications.

Spelman Miller, K. (2005). Second language writing research and pedagogy: A role for computer logging? *Computers and Composition, 22*, 297 - 317.

Spelman Miller, K., & Sullivan, K. (2006). Key-stroke logging: an introduction. In G. Rijlaarsdam (Series Ed.), K. Sullivan & E. Lindgren (Vol Eds.), *Studies in Writing Vol. 18, Computer key-stroke logging and writing: methods and applications* (pp. 1 - 9). Oxford: Elsevier.

Stevenson, M. (2005). *Reading and writing in a foreign language; a comparison of conceptual and linguistic processes in Dutch and English*. SCO-Kohnstamm Instituut van de Faculteit der Maatschappij- en Gedragswetenschappen, University of Amsterdam, Amsterdam.

Stevenson, M., Schoonen, R., & De Glopper, K. (2006). Revising in two languages: A multi-dimensional comparison of online writing revisions in L1 and FL. *Journal of Second Language Writing, 15*, 201-233.

Sullivan, K., & Lindgren, E. (Eds.). (2006). *Computer key-stroke logging and writing: methods and applications* (Vol. 18). Oxford: Elsevier.

Thorson, H. (2000). Using the computer to compare foreign and native language writing processes: a statistical and case study approach. *The Modern Language Journal, 84*(ii), 155 - 170.

Tillema, M., Van den Bergh, H., Rijlaarsdam, G., & Sanders, T. (2008). *Adaptive strategies during writing in L1 and FL*. Manuscript submitted for publication.

Torrance, M. (1996). Is writing expertise like other kinds of expertise? In G. Rijlaarsdam & E. Espéret (Series Ed.), G. Rijlaarsdam, H. Van den Bergh & M. Couzijn (Vol Eds.), *Studies in Writing Vol. 2, Theories, models and methodology in writing research* (pp. 3 - 9). Amsterdam: Amsterdam University Press.

Uzawa, K. (1996). Second language learners' processes of L1 writing, L2 writing and translation from L1 into L2. *Journal of Second Language Writing, 5*(3), 271-294.

Uzawa, K., & Cumming, A. (1989). Writing strategies in Japanese as a foreign language: lowering or keeping up the standards. *The Canadian Modern Language Review/La Revue canadienne des languages vivantes, 46*(1), 178 - 194.

Van den Bergh, H. (1999). Genereren en genereren is twee [Generating and generating makes two]. In R. Hamel, M. Elshout-Mohr & M. Milikowski (Eds.), *Meesterschap: 16 stukken over intelligentie, leren, denken & probleemoplossen voor Jan Elshout* (pp. 9 - 21). Amsterdam: Amsterdam University Press.

Van den Bergh, H., De Glopper, K., & Schoonen, R. (1988). Directe metingen van schrijfvaardigheid: validiteit en taakeffecten. [Direct measures of writing skill: validity and task effects] In F. H. Van Emeren & R. Grootendorst (Eds.), *Taalbeheersing in ontwikkeling* (pp. 370 - 378). Dordrecht, The Netherlands: Foris Publications.

Van den Bergh, H., Herrlitz, W., & Klein Gunnewiek, L. (1999). Entwicklung von Fähigkeiten in einer Fremdsprache Vorsichtige Hypothesen. In G. Kischel &

E. Gothsch (Eds.), *Wege zur Mehrsprachigkeit im Fernstudium* (pp. 115 - 138). Hagen: Domröse.

Van den Bergh, H., & Rijlaarsdam, G. (1996). The dynamics of composing: modelling writing process data. In C. M. Levy & S. Ransdell (Eds.), *The science of writing. Theories, methods, individual differences & applications* (pp. 207 - 232). Mahwah, NJ: Lawrence Erlbaum Associates.

Van den Bergh, H., & Rijlaarsdam, G. (1999). The dynamics of idea generation during writing: an on-line study. In G. Rijlaarsdam & E. Espéret (Series Eds.), M. Torrance & D. Galbraith (Vol Eds.), *Studies in Writing Vol. 4, Knowing what to write; Conceptual processes in text production* (pp. 99 - 120). Amsterdam: Amsterdam University Press.

Van den Bergh, H., & Rijlaarsdam, G. (2001). Changes in cognitive activities during the writing process and relationships with text quality. *Educational Psychology, 21*(4), 373-385.

Van den Bergh, H., & Rijlaarsdam, G. (2007). The dynamics of idea generation during writing: an online study. In G. Rijlaarsdam (Series Ed.), M. Torrance, L. Van Waes & D. Galbraith (Vol Eds.), *Studies in writing Vol. 20, Writing and cognition: research and applications* (pp. 125 - 150). Amsterdam: Elsevier.

Van den Bergh, H., Rijlaarsdam, G., Janssen, T., Braaksma, M., Van Weijen, D., & Tillema, M. (in press). Relating process execution to writing, reading and output quality of text: consideration learner and task characteristics. In M. C. Shelley II, L. D. Yore & B. Hand (Eds.), *Quality research in literacy and science education: International perspectives and gold standards* (pp. XXX - XXX). Dordrecht, The Netherlands: Springer.

Van der Hoeven, J. (1996a). Children's composing: a study of the relationships between planning and revision skill, writing processes, text quality and linguistic skills. In G. Rijlaarsdam & E. Espéret (Series Eds.), G. Rijlaarsdam, H. Van den Bergh & M. Couzijn (Vol Eds.), *Studies in writing Vol. 2, Theories, models and methodology in writing research* (pp. 108 - 120). Amsterdam: Amsterdam University Press.

Van der Hoeven, J. (1996b). *Children's composing: A study into the relationship between writing processes, text quality, and cognitive and linguistic skills.* Amsterdam: Rodopi.

Van der Hoeven, J. (1999). Differences in writing performance: Generating as indicator. In G. Rijlaarsdam & E. Espéret (Series Eds.), M. Torrance & D. Galbraith (Eds.), *Studies in writing Vol. 4, Knowing what to write; Conceptual processes in text production* (pp. 65-77). Amsterdam: Amsterdam University Press.

Van Steendam, E., Rijlaarsdam, G., & Sercu, L. (2007). *Improving text coherence: effects ofcollaborative revisions conditions on writing quality in EFL.* Paper presented at the Developing potentials for learning. Programma of the 12th Biennial Conference for Research on Learning and Instruction, Budapest, Hungary.

Van Steendam, E., Rijlaarsdam, G., & Sercu, L. (2008). *The effect of strategy instruction on the quality of peer feedback in ESL/EFL: Pair vs. individual learning, practising vs. observing.* Manuscript submitted for publication.

Van Waes, L., & Schellens, P. J. (2003). Writing Profiles: the effect of the writing mode on pausing and revision patterns of experienced writers. *Journal of Pragmatics, 35*(6), 829 - 853.

Van Weijen, D. (2006). Expanding the unit of analysis in writing process research: a promising first exploration. In C. Keskin (Ed.), *Yearbook Utrecht Institute of Linguistics* (pp. 147 - 161). Utrecht.

Van Weijen, D., Van den Bergh, H., Rijlaarsdam, G., & Sanders, T. (2005a). Writing processes in L1 and in a foreign language: a first exploration. In L. Allal & J. Dolz (Eds.), *Proceedings Writing 2004 [CD]* (Adcom Productions [available on-line: sigwritingpublications.org].

Van Weijen, D., Van den Bergh, H., Rijlaarsdam, G., & Sanders, T. (2005b). *Adaptivity: transferring writing processes between tasks (in L1) and between languages (from L1 to FL English).* Paper presented at the 11th European Conference for Learning and Instruction, Nicosia, Cyprus, August 22-27, 2005.

Wang, L. , (2003). Switching to first language among writers with differing second-language proficiency. *Journal of Second Language Writing, 12*(4).

Wang, W. , & Wen, Q. (2002). L1 use in the L2 composing process: an exploratory study of 16 Chinese EFL writers. *Journal of Second Language Writing, 11*(3), 225-246.

Wesdorp, H. (1974). Het meten van de 'stelvaardigheid' [Measuring 'writing ability']. *Pedagogische Studiën, 51*, 499-521.

Whalen, K., & Ménard, N. (1995). L1 and L2 writers' strategic and linguistic knowledge: a model of multiple-level discourse processing. *Language Learning, 45*(3), 381 - 418.

Wolfersberger, M. (2003). L1 to L2 writing process and strategy transfer: a look at lower proficiency writers. *TESL-EJ: Teaching English as a Second or Foreign Language, 7*(2), A6 1-15.

Woodall, B. R. (2002). Language-switching: using the first language while writing in a second. *Journal of Second Language Writing, 11*, 7 - 28.

Zimmermann, R. (2000). L2 writing: subprocesses, a model of formulating and empirical findings. *Learning and Instruction, 10*, 73 - 99.

AUTHOR INDEX

—S—

—T—

—U—

—V—

—W—

—Y—

—Z—

SAMENVATTING

Het schrijven van een tekst kost vaak veel moeite. Er wordt zelfs beweerd dat het schrijven van een tekst even veel energie kost als het graven van een greppel (Kellogg, 1994, p. 17). Dat schrijvers desondanks in staat zijn om teksten, en soms zelfs heel erg goede teksten, te produceren lijkt haast een wonder. Maar, hoe produceren schrijvers hun teksten, en nog belangrijker, hoe worden goede teksten geschreven? Die vraag is in dit proefschrift aan de orde.

Tijdens het schrijven maken schrijvers gebruik van verschillende cognitieve activiteiten, zoals het plannen van de tekst, het genereren van ideeën, het formuleren van de inhoud en het evalueren en reviseren van de tekst. Daarnaast lezen zij tijdens het schrijven regelmatig hun eigen tekst-in-wording en, indien aanwezig, soms ook de opdracht op basis waarvan ze aan het schrijven zijn. Uit onderzoek blijkt dat schrijvers verschillen in de manier waarop zij deze activiteiten inzetten tijdens het schrijven. Zo verschillen ze bijvoorbeeld in de mate waarin ze een bepaalde activiteit toepassen, of wanneer ze dat doen tijdens het schrijven. Sommige schrijvers beginnen bijvoorbeeld met het genereren van ideeën, terwijl andere schrijvers dit uitstellen en eerst nadenken over de opbouw of structuur van de tekst. Tot nu toe is in wetenschappelijk onderzoek echter relatief weinig aandacht besteed aan de vraag wat dergelijke verschillen tussen schrijvers betekenen voor de kwaliteit van de teksten die zij schrijven. Hoe beïnvloedt de manier waarop schrijvers te werk gaan tijdens het schrijven de kwaliteit van de teksten die zij produceren?

Uit eerder onderzoek blijkt dat de manier waarop schrijvers de verschillende activiteiten inzetten verandert tijdens het schrijfproces. Aan het begin van het schrijfproces wordt bijvoorbeeld veel aandacht besteed aan het lezen van de opdracht, terwijl de opdracht in latere fases van het schrijfproces een veel minder belangrijke rol speelt. Maar dit verschilt enorm tussen schrijvers. De ene schrijver herleest de opdracht regelmatig tijdens het schrijfproces, bijvoorbeeld om te controleren of zijn tekst nog voldoet aan de gestelde eisen, terwijl een andere schrijver het doel van de opdracht aan het einde van het schrijfproces nog duidelijk voor ogen heeft en de opdracht dus niet meer hoeft te herlezen. Daarnaast blijkt ook dat juist het moment waarop schrijvers verschillende activiteiten toepassen van belang is voor de kwaliteit van de tekst die geschreven wordt. Het maakt voor de kwaliteit van de tekst dus niet zozeer uit hoeveel ideeën een schrijver genereert, of hoeveel revisies hij uitvoert, maar wel wanneer hij dat doet.

Het doel van het onderzoek dat in dit proefschrift gepresenteerd wordt is om onze kennis over het schrijfproces te verdiepen door het verband te onderzoeken tussen het schrijfproces en de kwaliteit van het product, de tekst. Het onderzoek richt zich in het bijzonder op de invloed van één specifiek taakkenmerk,

de taal waarin geschreven wordt, op het schrijfproces én op proces-productrelaties. Daarom wordt in dit onderzoek een vergelijking gemaakt tussen de manier waarop schrijvers schrijven in hun moedertaal, Nederlands, en hun tweede taal, Engels. Die vergelijking vindt plaats aan de hand van twee vragen: *1. Hoe verloopt het schrijfproces in moedertaal en tweede taal?* en *2. Wat is de relatie tussen het verloop van het schrijfproces en tekstkwaliteit in moedertaal en tweede taal?*

Om deze vragen te beantwoorden, willen we op basis van onze resultaten conclusies kunnen trekken over het schrijfproces in het algemeen. Dat betekent dat we onze resultaten moeten kunnen generaliseren over taken. Dat is echter onmogelijk op basis van één schrijftaak per persoon. Als iedere proefpersoon slechts twee teksten schrijft (één in het Nederlands en één in het Engels), dan hoeven de verschillen tussen die twee schrijfprocessen niet te liggen aan de taal waarin geschreven wordt. Eventuele verschillen kunnen ook liggen aan het feit dat het simpelweg twee verschillende schrijftaken met twee verschillende onderwerpen zijn. Dat betekent dat er meer taken per persoon afgenomen moeten worden. Bovendien is een bijkomend probleem dat we op basis van één taak per schrijver niet kunnen bepalen of die schrijver een goede of juist zwakke schrijver is. We kunnen wel bepalen of zijn tekst een goede of slechte tekst is, maar we kunnen onmogelijk bepalen of die tekst kwalitatief vergelijkbaar is met teksten die hij op andere momenten zou hebben geschreven. Misschien is de kwaliteit van die ene tekst juist veel beter of slechter dan normaal, omdat de schrijver een goede of juist een slechte dag heeft. Daarom is het, zowel voor het bestuderen van het schrijfproces, als voor het bepalen van tekstkwaliteit, van groot belang dat schrijvers in dit onderzoek meerdere teksten schrijven in beide talen.

Aan dit onderzoek deden 20 eerstejaarsstudenten Engels mee. Zij kregen de opdracht om acht korte betogende teksten te schrijven, vier in hun moedertaal, Nederlands, en vier in hun tweede taal, Engels. Daarnaast kregen deze proefpersonen eerst een korte training om te oefenen met het schrijven onder hardopdenkcondities. Zij moesten proberen om hun gedachten zoveel mogelijk hardop onder woorden te brengen tijdens het schrijven van hun teksten. Daardoor werd het mogelijk om te bepalen welke cognitieve activiteiten zij gebruikten tijdens het schrijven en wanneer ze dat deden. Proefpersonen werden geobserveerd tijdens het schrijven en er werden audio- en video-opnames gemaakt van elke schrijfsessie. Dit resulteerde in 80 Nederlandse en 79 Engelse hardopdenkprotocollen.

Elk protocol is een weergave van het schrijfproces van één taak van één schrijver en bevat informatie over de tekst-in-wording en over de gedachten die daaraan ten grondslag liggen. Daarnaast bestaat elk protocol uit een groot aantal protocolfragmenten, maar elk fragment bevat slechts één cognitieve activiteit. Dat betekent dat als er van een cognitieve activiteit overgestapt werd op een andere activiteit (bijvoorbeeld van plannen naar formuleren), dan werd die nieuwe activiteit in een nieuw fragment gezet. Vervolgens zijn deze hardopdenkprotocollen op

verschillende manieren geanalyseerd om het effect te bepalen van schrijven in moedertaal en tweede taal op het verloop van het schrijfproces en op de kwaliteit van de geschreven teksten.

RESULTATEN

Schrijven in de moedertaal en een tweede taal

Het belangrijkste doel van dit onderzoeksproject was om te bepalen wat de invloed is van de taal waarin geschreven wordt (moedertaal of tweede taal) op het gebruik van verschillende cognitieve activiteiten tijdens het schrijfproces en op de kwaliteit van de geproduceerde teksten. We hebben ons eerst geconcentreerd op het schrijven van Nederlandse teksten (zie hoofdstuk 2), en daarna op het schrijven van Engelse teksten (zie hoofdstuk 3). Voor beide talen hebben we onderzocht hoe schrijvers gebruik maken van de verschillende cognitieve activiteiten tijdens het schrijven én wat de invloed daarvan is op de kwaliteit van de geproduceerde teksten. Pas daarna is een vergelijking gemaakt tussen schrijven in het Nederlands en in het Engels (zie hoofdstuk 4). Bij deze vergelijking staan zowel verschillen in het schrijfproces als verschillen in proces-productrelaties centraal. Daarbij hebben we vooral gekeken naar de rol van het lezen van de opdracht, plannen, het genereren van ideeën en formuleren tijdens het schrijven in beide talen.

Uit de resultaten blijkt dat het moment waarop deze activiteiten voorkomen tijdens het schrijfproces verschilt tussen schrijvers. De ene schrijver begint bijvoorbeeld eerder (of juist later) met het genereren van ideeën dan de ander. Daarnaast verschillen schrijvers in de mate waarin zij het gebruik van deze activiteiten variëren tussen taken. Sommige schrijvers pakken al hun schrijftaken op een vergelijkbare manier aan, terwijl andere schrijvers de manier waarop zij te werk gaan deels veranderen van taak tot taak. Verder verschilt het moment waarop de cognitieve activiteiten een positieve invloed lijken te hebben op tekstkwaliteit. Zo blijkt het genereren van ideeën vooral positief gerelateerd aan tekstkwaliteit aan het begin van het schrijfproces, terwijl die relatie voor formuleren juist positief blijkt te zijn in het midden van het schrijfproces.

Bij een vergelijking van schrijven in de moedertaal- en in een tweede taal, blijken er significante verschillen te zijn in de momenten waarop de cognitieve activiteiten toegepast worden gedurende het schrijfproces in beide talen. Maar een kwalitatieve vergelijking levert ook overeenkomsten op (zie hoofdstuk 4). Zo blijkt dat schrijvers in beide talen aan het begin van het schrijfproces vooral aandacht besteden aan het lezen van de opdracht, terwijl zij in het midden van het schrijfproces vooral bezig zijn met formuleren. Daarnaast wordt in beide talen relatief gezien veel meer aandacht besteed aan formuleren dan aan plannen of het genereren van ideeën.

Een ander verschil tussen schrijven in beide talen is gerelateerd aan het moment waarop het toepassen van een activiteit een positieve invloed heeft op tekstkwaliteit. Het lezen van de opdracht levert bijvoorbeeld in beide talen hoogstwaarschijnlijk een positieve bijdrage aan tekstkwaliteit als het gebeurt aan het begin van het schrijfproces. In latere fases van het schrijfproces lijkt het herlezen van de opdracht echter een negatieve invloed te hebben op tekstkwaliteit in het Engels, terwijl het in het Nederlands weinig uitmaakt of schrijvers de opdracht later nog herlezen of niet.

Het meest opvallende verschil tussen het schrijven in beide talen is het feit dat schrijvers het moment waarop zij de verschillende activiteiten toepassen veel minder variëren van taak tot taak tijdens het schrijven in hun tweede taal, dan in hun moedertaal. Met andere woorden, het gedrag van schrijvers blijkt gemiddeld veel consistenter of stabieler tussen schrijftaken tijdens het schrijven in een tweede taal dan in de moedertaal. Het verschil in mate van variatie tussen taken lijkt dus veroorzaakt te worden door de taal waarin geschreven wordt. Maar dat is eigenlijk een te algemene verklaring. Specifieker gezegd zouden de verschillen te maken kunnen hebben met het feit dat schrijven in een tweede taal over het algemeen een nog complexere taak is dan schrijven in je moedertaal. Op procesniveau zou het daarom zo kunnen zijn dat sommige schrijvers een zekere mate van cognitieve overbelasting ervaren tijdens het schrijven in een tweede taal. Dat zou betekenen dat ze, cognitief gezien, tijdens het schrijven in hun tweede taal minder ruimte beschikbaar hadden om hun aanpak aan te passen aan de specifieke eisen van de taak die ze moesten uitvoeren dan in hun moedertaal.

Interactie tussen activiteiten

We weten nu dat schrijvers hun aanpak van schrijftaken in hun moedertaal meer variëren dan in een tweede taal. Maar daarbij is alleen gekeken naar de rol van individuele activiteiten tijdens het schrijfproces. De onderlinge relaties tussen deze activiteiten zijn nog niet bestudeerd. Een belangrijke nieuwe bijdrage van dit proefschrift wordt dus gevormd door het onderzoek naar de vraag hoe de diverse activiteiten elkaar beïnvloeden tijdens het schrijven. Activiteiten zoals plannen of evalueren zouden bijvoorbeeld invloed kunnen uitoefenen op de aansturing van het schrijfproces als geheel. Dat zou kunnen betekenen dat het schrijven plaats vindt volgens een hiërarchische structuur, waarbij sommige activiteiten andere activiteiten aansturen. Om dit te onderzoeken hebben we een analyse uitgevoerd waarin we gebruik hebben gemaakt van *Composing Episodes*, ofwel schrijfepisodes (zie hoofdstuk 5). Deze schrijfepisodes bevatten ieder meerdere protocolfragmenten, waarin verschillende cognitieve activiteiten voorkomen. Het idee was dat activiteiten die aan het begin van episodes voorkwamen misschien een sturende functie zouden hebben op de activiteiten die later binnen die episodes voorkwamen.

Uit de analyse blijkt dat schrijvers gemiddeld minder schrijfepisodes gebruiken tijdens het schrijven in het Nederlands dan in het Engels. Daarnaast blijkt dat alleen plannen een positieve samenhang vertoont met tekstkwaliteit wanneer het voorkomt aan het begin van episodes. Dat zou kunnen betekenen dat plannen op het juiste moment, aan het begin van episodes, tekstkwaliteit kan beïnvloeden, doordat het invloed heeft op het schrijfproces als geheel. Dat lijkt ook een bevestiging te zijn voor het idee dat schrijven een hiërarchisch gestructureerd proces is.

Het gebruik van de moedertaal tijdens schrijven in een tweede taal

Uit eerder onderzoek is gebleken dat schrijvers tijdens het schrijven in een tweede taal soms gebruik maken van hun moedertaal. Dit geldt ook voor onze proefpersonen. Daarom is nagegaan of dit gerelateerd is aan de kwaliteit van de geproduceerde teksten. Zijn schrijvers die in het Engels denken en doen beter dan schrijvers die ook in het Nederlands denken of zelfs stukjes schrijven in het Nederlands en dan vertalen naar het Engels? Om dit te bepalen is onderzocht in hoeverre verschillende conceptuele activiteiten, zoals plannen en evalueren, voorkwamen in het Nederlands tijdens het schrijven in het Engels. Daarnaast hebben we gekeken of de mate waarin schrijvers hun moedertaal gebruikten gerelateerd was aan hun schrijfvaardigheid in het Nederlands, hun taalvaardigheid in het Engels en de kwaliteit van hun Engelse teksten.

Uit de analyses blijkt dat er een negatieve samenhang is tussen het gebruik van het Nederlands, van schrijfvaardigheid in het Nederlands en van taalvaardigheid in het Engels. Dat betekent dat vooral schrijvers die slechte schrijvers zijn in het Nederlands, en schrijvers met een lagere taalvaardigheid in het Engels vaak gebruik maken van hun moedertaal tijdens het schrijven in het Engels. Ook blijkt er een negatieve relatie te zijn tussen het gebruik van het Nederlands en tekstkwaliteit. Deze relatie wordt beïnvloed door zowel de Nederlandse schrijfvaardigheid van de schrijvers als door hun Engelse taalvaardigheid. Zo heeft het Nederlands alleen een negatieve invloed op de kwaliteit van de Engelse teksten als het gebruikt wordt door schrijvers met een hoge Engelse taalvaardigheid of door schrijvers met een lage schrijfvaardigheid in het Nederlands.

Verder is het gebruik van het Nederlands over het algemeen iets pregnanter aan het begin van het schrijfproces, maar er is slechts sprake van een zeer geringe afname gedurende het schrijfproces. Het moment waarop het Nederlands gebruikt wordt tijdens het schrijven in het Engels is dus waarschijnlijk niet van groot belang. Tot slot is het aannemelijk dat het gebruik van het Nederlands niet de oorzaak is van een daling in de kwaliteit van de Engelse teksten. Het lijkt eerder zo te zijn dat het feit dat schrijvers gebruik maken van hun moedertaal een teken is dat zij proberen om cognitieve overbelasting te voorkomen (of zich daarvan te bevrijden) tijdens het schrijven in hun tweede taal. Als die

poging vervolgens mislukt, dan zou de kwaliteit van de geschreven tekst daar onder kunnen lijden (zie hoofdstuk 6).

ALGEMENE CONCLUSIES

Op basis van de resultaten van dit proefschrift concludeer ik dat de relatie tussen het schrijfproces en de geschreven tekst op een aantal verschillende manieren verandert als schrijvers overstappen van hun moedertaal naar hun tweede taal. Allereerst blijkt uit onze analyses dat schrijven in een tweede taal het moment beïnvloedt waarop de verschillende cognitieve activiteiten optreden tijdens het schrijfproces. Daarnaast blijkt dat de positieve invloed die deze activiteiten hebben op de kwaliteit van de geschreven tekst afhangt van het moment waarop ze plaats vinden én van de taal waarin geschreven wordt. Het genereren van ideeën blijkt bijvoorbeeld in het Nederlands alleen een positieve invloed te hebben op tekstkwaliteit aan het begin van het schrijfproces, terwijl het in het Engels een positieve bijdrage levert tijdens het hele schrijfproces. Tot slot variëren schrijvers de manier waarop zij de verschillende activiteiten toepassen gemiddeld minder tussen taken tijdens het schrijven in een tweede taal dan tijdens het schrijven in hun moedertaal. Met andere woorden, als schrijvers schrijven in hun tweede taal, lijkt hun gedrag stabieler tussen taken dan wanneer ze schrijven in hun moedertaal. Dat zou kunnen liggen aan het feit dat zij het schrijven in hun tweede taal ervaren als een cognitief zwaardere taak dan schrijven in hun moedertaal. Of dat echt zo is zal in de toekomst nog nader worden onderzocht.

CURRICULUM VITAE

Daphne van Weijen was born on March 27th 1978 in Soest, The Netherlands. She attended the *International School Hilversum 'Alberdingk Thijm'* in Hilversum, where she obtained her International Baccalaureate diploma in 1996. In 2002 she graduated with distinction (cum laude) from Utrecht University, with a Liberal Arts MA (Algemene Letteren), specializing in Sociolinguistics and Language teaching. During her MA she spent half a year at the University of Sheffield (U.K.) in 2001, as an assistant Dutch teacher in the department of Germanic Studies. In 2004 she started her PhD research at the Utrecht Institute of Linguistics – OTS. This thesis is the result of the research she carried out during that period.