Differences in process and process-product relations in L2 writing
van Weijen, D.; van den Bergh, H.; Rijlaarsdam, G.C.W.; Sanders, T.

Published in:
ITL : International Journal of Applied Linguistics

DOI:
10.2143/ITL.156.0.2034433

Citation for published version (APA):

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

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
Differences in Process and Process-Product Relations in L2 Writing

Daphne van Weijen
UiL-OTS, Utrecht University
Huub van den Bergh
UiL-OTS, Utrecht University; ILO University of Amsterdam
Gert Rijlaarsdam
ILO, University of Amsterdam
Ted Sanders
UiL-OTS, Utrecht University

Abstract

This study examines 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 first year students wrote four texts each in their L2 (English) under think-aloud conditions. The analysis focused on four cognitive activities: Reading the assignment and sources, Planning, Generating ideas and Formulating. Results indicate that, on average, the occurrence of each 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 activities) while writing. These differences, however, depend on the moment in the writing process, and on the activity being carried out. In general, writers’ behaviour is rather stable between tasks, at least at the start of the writing process. Finally, results indicate that the correlation between each activity and text quality varies during the writing process and also differs somewhat between tasks.
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 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 vary 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 Van Weijen, Van den Bergh, Rijlaarsdam & Sanders, 2008a). In other words, the moment at which activities are carried out is essential when assessing process-product relations in L1 writing (see Rijlaarsdam & Van den Bergh, 2006, p. 42). For example, Van Weijen, et al. (2008a) found that for L1 writing, 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. Formulating, on the other hand, correlated positively with text quality when it occurred after the start of the writing process (see Van Weijen et al., 2008a). Evidently, changes in the 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 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. Earlier research on L2 writing focused on: comparing L1 versus L2 writing (e.g., Beare & Bourdages, 2007; Chenoweth & Hayes, 2001; Roca de Larios, Manchón & Murphy, 2006; Roca de Larios, Marín & Murphy, 2001; Thorson, 2000), comparing different groups of L2 writers (e.g., Sasaki, 2002, 2004; Shi, 2004; see Cumming, 2001, p. 5 for an overview), the use of L1 during L2 writing (e.g., Beare & Bourdages, 2007; Qi, 1998; Wang & Wen, 2002; Woodall, 2002), or on building L2 writing models (e.g., Cumming, 2001; Grabe, 2001; Roca de Larios, Murphy & Marín, 2002; Silva, 1993; Wang & Wen, 2002). However, research on the relation between the writing process (i.e., the cognitive activities employed during writing) and text quality in L2 was somewhat scarce.

The few studies on process-product relations in L2, focused on a single cognitive activity instead of on 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 process characteristics such as fluency or metacognitive knowledge on text quality (e.g., Schoonen, Van Gelderen, DeGlopper, Hulstijn, Simis, Snellings, et al., 2003; Stevenson, 2005), or concentrated on relating the writing process to product characteristics, such as accuracy and syntactic complexity (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. Such a single task design hinders generalizations, because their data provide no indications of variation between tasks within each language. Furthermore, most studies that compare 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). Hence, time will be included as a factor in this study, with the following main question: Are changes in task representation reflected in the writing process in L2 writing?

In this study, the analysis of process execution focuses on four cognitive activities. We chose to include two conceptual activities, Planning and Generating, which are rather language independent, and two linguistic activities, Reading the assignment and Formulating, which are clearly language dependent (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 a factor that possibly influences a writer’s initial task representation (see Kennedy, 1985). Furthermore it is especially important in educational settings. Planning was included because it is one of the most important elements in the writing process (see for instance Flower & Hayes, 1980, 1981; Breetvelt et al., 1994; Levy & Ransdell, 1995; Van den Bergh & Rijlaarsdam, 2001) as it appears to govern the writing process as a whole. We included Generating because ideas are a prerequisite for text production (see 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 will vary during the writing process in L2. 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. Some writers will concentrate more on Reading the assignment in the beginning of the writing process for example, 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. For example, writers will not focus on Generating at more or less the same time during each task. Fourth, the relation between the occurrence of each activity and text quality will vary during the writing process in L2. 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%\(^1\) and their average age was 18 years and 10 months. 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. 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 and consisted of an algebra puzzle, a crossword puzzle and a short writing assignment. They had to write approximately five lines on one of ten randomly assigned topics, such as their favourite music or book. To re-familiarize them with thinking aloud, they had to write a similar text, at the start of each subsequent task.

The use of the think-aloud method in writing research is problematic (see for example Ericsson, 1998; Roca, de Larios, et al., 2006). But Janssen, Van Waes & Van den Bergh (1996) reported 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 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 was no evidence that fundamental production processes proceeded in a qualitatively different way in both conditions. Thus, 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

---

\(^1\) The male-female ratio of the participants in this study is a reflection of the male-female ratio of the student population at the institute where the study was carried out.
One participant only wrote three essays instead of four due to technical difficulties.

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; Levy & Ransdell, 1995; Couzijn, et al., 2002; Roca de Larios, et al 2006; Stevenson, 2005; Uzawa, 1996; Wang & Wen, 2002). Therefore, 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

Participants each wrote four argumentative essays for a fictitious essay contest in the University Newspaper, on four out of eight topics included in the study (see Table 3). The topics were randomly distributed over participants in a balanced design. 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, goal and target audience) as well as six concise sources, provided to give participants some background information on the topic. They had to quote at least two of the sources in each essay.

Data

After collecting the data, we transcribed the audio and video recordings of each individual session to form think-aloud protocols (N=79)². The protocols were segmented, so that each 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).

Table 1: Overview of the cognitive activities included in the coding scheme

<table>
<thead>
<tr>
<th>Activity</th>
<th>Description and Protocol Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Re)-Reading the assignment</td>
<td>Reading the assignment and sources</td>
</tr>
<tr>
<td>Planning: Self-instructions</td>
<td>Instructions the participant gives himself concerning the writing process (e.g., “I need to decide how to start”)</td>
</tr>
<tr>
<td>Planning: Goal-setting</td>
<td>Paraphrasing task demands or setting new goals (e.g., “This text is going to be about...”)</td>
</tr>
<tr>
<td>Planning: Structuring</td>
<td>Selecting and ordering ideas or making a rough outline (e.g., “First the pro’s, then the con’s and then my own opinion!”)</td>
</tr>
<tr>
<td>Generating ideas</td>
<td>Paraphrasing ideas from the assignment or generating new ones: (e.g., “Us humans we’re very curious and we like to gossip!”)</td>
</tr>
<tr>
<td>Metacommets</td>
<td>Comments on the writing process, assignment, or sources (e.g., “I had the perfect sentence in my head”)</td>
</tr>
<tr>
<td>Pausing</td>
<td>Silence or interjections (e.g., um, ehhh)</td>
</tr>
</tbody>
</table>

² One participant only wrote three essays instead of four due to technical difficulties.
Descriptive empirical studies

<table>
<thead>
<tr>
<th>Cognitive Activity</th>
<th>Average number of segments (sd)</th>
<th>Range of number of segments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reading the assignment</td>
<td>37.20 (13.81)</td>
<td>Min: 12, Max: 88</td>
</tr>
<tr>
<td>Planning: total</td>
<td>25.01 (19.59)</td>
<td>Min: 0, Max: 91</td>
</tr>
<tr>
<td>Planning: self-instructions</td>
<td>23.13 (17.83)</td>
<td>Min: 0, Max: 84</td>
</tr>
<tr>
<td>Planning: goal-setting</td>
<td>0.76 (2.54)</td>
<td>Min: 0, Max: 19</td>
</tr>
<tr>
<td>Planning: structuring</td>
<td>1.20 (2.21)</td>
<td>Min: 0, Max: 13</td>
</tr>
<tr>
<td>Generating ideas</td>
<td>41.51 (40.11)</td>
<td>Min: 1, Max: 215</td>
</tr>
<tr>
<td>Formulating</td>
<td>134.72 (39.69)</td>
<td>Min: 54, Max: 243</td>
</tr>
<tr>
<td>Writing process as a whole</td>
<td>875.73 (269.92)</td>
<td>Min: 390, Max: 1653</td>
</tr>
</tbody>
</table>

Subsequently, the protocols were coded using a scheme based on Hayes and Flower (1980) and frequently used in earlier studies (e.g., Breetvelt, et al., 1994; Couzijn, et al., 2002; Van den Bergh & Rijlaarsdam, 1999; Van der Hoeven, 1996; Van Weijen, et al., 2008a). The activities included in the coding scheme are shown in Table 1. Research on L1 writing has shown that these 11 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).

In this study, we defined the activities included in the analysis as follows. Reading the assignment was defined as reading both the assignment and 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). We did not include Generating ideas in the planning category, 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).

Table 2: Average frequency of occurrence and range for each cognitive activity averaged over all texts and writers (N = 79)
This study focuses 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 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 shows that Formulating is, on average, the most frequent process, occurring 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 type 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 a jury of five judges: analytically and holistically. The 79 texts were first rated using an analytical rating form, with four main categories: Structure, Content, Argumentation and Conclusion. We modified the original categories in an earlier form (see Breetvelt et al., 1994) so that the focus was on structure and argumentation (see Sanders & Schilperoord, 2006), as these are likely to be stable features in texts written by the same writer, possibly even between languages.

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

<table>
<thead>
<tr>
<th>Task</th>
<th>Topic</th>
<th>N</th>
<th>Mean (sd)</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Education in English: good idea or bad suggestion?</td>
<td>9</td>
<td>-.15 (1.31)</td>
<td>-1.96</td>
<td>1.72</td>
</tr>
<tr>
<td>B</td>
<td>Having children, yes or no?</td>
<td>10</td>
<td>.17 (0.97)</td>
<td>-1.32</td>
<td>1.68</td>
</tr>
<tr>
<td>C</td>
<td>Everybody should automatically be a donor, unless they explicitly register against it: good or bad idea?</td>
<td>10</td>
<td>.19 (1.09)</td>
<td>-1.49</td>
<td>1.82</td>
</tr>
<tr>
<td>D</td>
<td>Life as a student: A hard life full of stress? Or isn’t it that bad?</td>
<td>10</td>
<td>.31 (0.45)</td>
<td>-.55</td>
<td>0.80</td>
</tr>
<tr>
<td>E</td>
<td>Do surveillance cameras in inner city areas increase public security?</td>
<td>10</td>
<td>.15 (1.23)</td>
<td>-2.36</td>
<td>1.54</td>
</tr>
<tr>
<td>F</td>
<td>Mobile phones: irritating or essential?</td>
<td>10</td>
<td>-.32 (1.34)</td>
<td>-3.08</td>
<td>1.53</td>
</tr>
<tr>
<td>G</td>
<td>Downloading Music for Free: criminal or should be possible?</td>
<td>10</td>
<td>-.17 (0.53)</td>
<td>-1.37</td>
<td>0.66</td>
</tr>
<tr>
<td>H</td>
<td>Soft drugs: should they be legalised or prohibited completely?</td>
<td>10</td>
<td>.18 (0.89)</td>
<td>-1.46</td>
<td>1.77</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>79</td>
<td>0.05 (1)</td>
<td>-3.08</td>
<td>1.82</td>
</tr>
</tbody>
</table>
Descriptive empirical studies

The holistic assessment was carried out two weeks later. Judges were 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 scores 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 contains the segment level ($n = 69182$). Therefore, we performed a multilevel analysis, which enables us to do justice to this hierarchy. Furthermore, multilevel analysis enables us to model the occurrence of each cognitive activity during the writing process. For this analysis, we chose 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 $i$ ($i = 1, 2, \ldots, i_p$), during task $j$ ($j = 1, 2, 3, 4$), for writer $k$ ($k = 1, 2, \ldots, 20$) during the writing process. Therefore, we can write the model as:

$$Logit(Y_{ijk}) = \beta_{0jk} \times M_{0ijk} + \beta_{1jk} \times M_{1ijk} + \beta_{2jk} \times M_{2ijk} + \ldots$$  \hspace{1cm} (1)

The polynomial model can be adapted to include as many powers of the moment (M0, M1, M2, ...) as are necessary to describe the data. 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:

$$\beta_{0jk} = \beta_{000} + u_{0jk} + v_{00k}$$
$$\beta_{1jk} = \beta_{100} + u_{1jk} + v_{10k}$$
$$\ldots$$  \hspace{1cm} (2)

3 Please note: logit ($P$) = ln ($\frac{e^P}{1+e^P}$)
4 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 into three equal parts, as in earlier research (e.g., Breetvelt et al., 1994).
This means that the variance between tasks ($S^2 u_{tjk}$, $S^2 u_{tjk}$, ...) and individuals ($S^2 \beta_{00k}$, $S^2 \beta_{00k}$, ...) are estimated for each of the mean regression weights ($\beta_{00k}$, $\beta_{00k}$, ...). 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 ($v_{0jk}$, $v_{10k}$, ...). Furthermore, 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 ($u_{tjk}$, $u_{tjk}$, ...). If we insert Equation 2 in Equation 1, we obtain the model to be estimated:

\[
\text{Logit}(Y_{ijk}) = M^0_{jk}(\beta_{00k} + u_{0jk} + v_{00k}) + \\
M^1_{jk}(\beta_{100} + u_{1jk} + v_{10k}) + \\
M^2_{jk}(\beta_{200} + u_{2jk} + v_{20k}) + \\
\text{...}
\]

The model, as shown in Equation 3, has a fixed and a random part. The fixed part ($\beta_{00k}$, $\beta_{100}$, ...) 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 $M_k$ can be estimated as:

\[
S^2_{task} = S^2_{u0jk} + 2 \cdot S^2_{u0jk, u1jk} \cdot M^1_{jk} + S^2_{u1jk} \cdot M^2_{jk} + ...
\]

We can use the same type of equation to note the variance between individuals, by replacing the residuals for tasks ($u_{tjk}$, $u_{tjk}$, ...) with those that represent the deviations for individual $j$ ($v_{tjk}$, $v_{tjk}$, ...).

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 $j$ ($u_{tjk}$, $u_{tjk}$, ...) and the residuals for individuals ($v_{tjk}$, $v_{tjk}$, ...). It is important to remember that these correlations depend on the value of $M_k$ and its subsequent powers (see Van den Bergh & Rijlaarsdam, 1996).

To answer the main question, we must first determine the average curve that depicts the temporal distribution for each cognitive activity in L2. This is the fixed part of the model. Next, we must determine whether there is variation in process execution in L2 writing at all, and if so, whether there is mainly inter- or intra-writer variation or both. These questions pertain to the random part of the model. In the final step, we will determine the correlation between the occurrence of each cognitive activity and text quality at each moment in time (see Van Weijen et al., 2008a; Van den Bergh & Rijlaarsdam, 1996).
Descriptive empirical studies

Results

The parameter estimates of the model are presented in logits^5, a nonlinear transformation of proportions (see Equations 1 – 3 above). As this makes them hard to interpret, they were converted to proportions, indicating how likely an activity is to occur at specific points during the writing process. Please note that the curves in some of the figures differ in length. Participants did not spend the same amount of time on each task, which means that the activities and thus the curves varied in length. 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 after segment 1200 are based on observations of only a few writers and should thus be interpreted with caution.

Average temporal distributions

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. Furthermore, the temporal distribution of each activity is different. Reading is most likely to occur, on average at the start and end of the writing process, while Formulating is most likely to occur in the middle. Finally, Planning and Generating are most likely to occur in the beginning. Thus, the occurrence of activities appears to vary over time.

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 shows the average probability of occurrence per writer (averaged over four tasks) for each activity (n = 20). This indicates that the patterns differ between writers. For Reading, the differences between writers are large initially, but then decrease over time. For Planning and Generating, the probability of occurrence is generally low and differences between writers remain rather constant throughout the writing process, with one or two exceptions. 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 writers’ formulating behaviour gradually decreases over time, with one or two exceptions.
Figure 2: The average probability of occurrence for each cognitive activity per writer (n = 20)

Variation within writers

Next we examined the variation in the occurrence of each activity between tasks (within writers). Therefore, we calculated the average probability of occurrence for each activity per writer and the 80% confidence intervals which indicate how much a writer’s behaviour for a specific activity changes between tasks. The larger the interval, the more a writer’s behaviour is likely to vary strongly 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). The letters on the curves (A – H) on the right hand side of the figure, refer to the topic of the task (see Table 3).

The intrawriter variation for Reading is presented in Figure 3 (top row). On average, the probability of occurrence for Reading varies both within writers and over time. 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. The extent to which she reads the assignment varies between tasks, 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. Hence, it seems that the intrawriter variation is largely due to her Reading behaviour during Task F.

In general, Planning is a rather infrequent activity. Figure 3 shows, however, that its probability of occurrence varies between tasks (see Figure 3, second row, left). The within writer differences 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
Descriptive empirical studies

writer; only Task H differs markedly from the others. However, the probability of occurrence for Tasks C and H decreases gradually over time, while for Task G it increases rather sharply. If a writer encounters problems while writing, he is likely to produce self-instructions, a type of planning, in an attempt to solve these problems.

The intrawriter variation for Generating appears to be fairly constant, on average, as the distances between the curves are more or less stable throughout the writing process. The results for Writer 1 (see Figure 3, third row, right), however, show that this general trend 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 far more likely to generate new ideas, than during her other three tasks.

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))
Finally, for Formulating, the intrawriter variation is highest, on average, in the middle, 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). Her between-task-variation is rather small initially, but the differences between tasks increase slowly and become largest in the middle of the writing process.

**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 proportion of 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 therefore also indicates how different tasks affect the writing process.

The results of this analysis are presented in Table 4.

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

<table>
<thead>
<tr>
<th>Activity</th>
<th>Start</th>
<th>Middle</th>
<th>End</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reading the assignment</td>
<td>.75</td>
<td>.62</td>
<td>.65</td>
</tr>
<tr>
<td>Planning</td>
<td>.88</td>
<td>.61</td>
<td>.45</td>
</tr>
<tr>
<td>Generating</td>
<td>.71</td>
<td>.70</td>
<td>.58</td>
</tr>
<tr>
<td>Formulating</td>
<td>.77</td>
<td>.33</td>
<td>.73</td>
</tr>
</tbody>
</table>

Overall the proportion of variation due to individuals is much larger than the proportion of variation due to tasks. Writers’ behaviour is generally rather stable between tasks, although the proportion of variation due to individuals does vary over time. For all activities it is largest at the start of the writing process, and then appears to decrease over time, except for Formulating, for which it is larger near the end of the writing process than in the middle. So 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. 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, particularly at the start of the writing process.
Variation and text quality

Subsequently the correlation between the occurrence of each activity and text quality was estimated. 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, after which we will discuss how the correlation between each activity and text quality varies between tasks.

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

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 Planning, the correlation with text quality (Figure 4, top right) is negative initially, 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, and then gradually decreases over time. In general, Generating ideas seems to be a useful thing to do throughout the writing process. For Formulating, the correlation with text quality is negative at first, but increases over time, becoming positive shortly after the start of the writing process. Thus, students who are likely to focus on Formulating after the start of the writing process are likely to have written better texts than students who are more 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, but the variation becomes much larger from the middle of the writing process onwards. 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. Reading only has a positive influence on text quality at the very start of the writing process. 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. Thus we would expect Writer 6 (see Figure 3) to have produced her best text during task F, but this was not the case. This could be due to her very high probability of occurrence for Reading at the end of the writing process for that task, which might have negated the positive effect of Reading in the beginning. The correlation between Planning and text quality varies somewhat initially, but the variation is largest near the end of the writing process. 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. 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 holds for example, for Writer 9 (see Figure 3). Based on the relatively large amount of planning he does during Task G, compared to his other three tasks, we would predict that this resulted in his highest quality text. But task G actually resulted in the poorest quality text this writer produced. Therefore, it seems likely that he experienced problems while writing and tried to resolve them by using self-instructions. Although the correlation between Generating and text quality is consistently positive, Figure 4 shows that it varies strongly between tasks over time. Thus 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 a higher quality text. This is confirmed by the results for Writer 1 (see Figure 3). She wrote her best text during Task H, when she was far more likely to generate ideas than during her other three tasks, and her poorest quality text during Task C, when she hardly generated ideas at all. For Formulating, finally, the variation due to task is relatively small initially, but increases
strongly over time, becoming largest near the end of the writing process. 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 found 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 (Van Weijen, et al., 2008a).

Second, the temporal distribution of each 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, while for Formulating the variation is rather large throughout the writing process. For Planning and Generating the variation between writers can be attributed to the behaviour of only a few writers.

Third, results show that the pattern of occurrence of each activity also varies within writers, between tasks. 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, the variation due to task is largest at the start and end of the writing process, for Formulating it is slightly higher in the middle, and for Planning it is largest near the end. But 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 in the beginning of the writing process. Indeed, process execution seems surprisingly stable within writers in L2, far more so than in L1 (see Van Weijen, et al., 2008a). Thus we can conclude that process execution varies both between and within writers, although it varies under influence of the cognitive activity employed as well as over time.
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 it is positive initially, but then becomes negative, while for Generating it is positive throughout the writing process. For Formulating and Planning on the other hand, it is negative at first, but becomes positive from the middle of the writing process onwards. For 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 L2 writing process execution under influence of the changing task situation and, if so, whether variation in process executions are related to variations in text quality. Based on our results, we can conclude that variation in process execution appears to be related to text quality. Writers appear to have a preference for a specific orchestration of cognitive activities during the L2 writing process; they seldom deviate from their default approach to the writing task. However, our analyses focused on only four cognitive activities. Whether our results also hold for other 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 activity and text quality varied under influence of the task at hand. So, variations in the way 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 to 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 Van Weijen, Van den Bergh, Rijlaarsdam & Sanders, 2008b). Thus 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 issue will be assessed experimentally in future research.


REFERENCES


Estimates and standard errors (s.e.) in logits for all components of the models for all four cognitive activities\(^7\)

**Fixed part of the model**

<table>
<thead>
<tr>
<th>Fixed part components</th>
<th>Reading the assignment estimate (s.e.)</th>
<th>Planning estimate (s.e.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\beta_0 \cdot M_{ijk})</td>
<td>-5.207 (0.062)</td>
<td>-3.571 (0.138)</td>
</tr>
<tr>
<td>(\beta_1 \cdot M_{ijk}^1)</td>
<td>-0.501 (0.000)</td>
<td>-0.034 (0.021)</td>
</tr>
<tr>
<td>(\beta_2 \cdot M_{ijk}^2)</td>
<td>0.096 (0.000)</td>
<td>---</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Fixed part components</th>
<th>Generating estimate (s.e.)</th>
<th>Formulating estimate (s.e.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\beta_0 \cdot M_{ijk})</td>
<td>-3.110 (0.155)</td>
<td>-1.487 (0.070)</td>
</tr>
<tr>
<td>(\beta_1 \cdot M_{ijk}^1)</td>
<td>-0.038 (0.017)</td>
<td>0.047 (0.017)</td>
</tr>
<tr>
<td>(\beta_2 \cdot M_{ijk}^2)</td>
<td>---</td>
<td>-0.023 (0.001)</td>
</tr>
</tbody>
</table>

**Random part: Variance between individuals**

<table>
<thead>
<tr>
<th>Random part components</th>
<th>Reading the assignment estimate</th>
<th>Planning estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>(S^2 \beta_0)</td>
<td>1.694</td>
<td>0.318</td>
</tr>
<tr>
<td>(S^2 \beta_1)</td>
<td>0.414</td>
<td>0.104</td>
</tr>
<tr>
<td>(S^2) Text quality*</td>
<td>-0.289</td>
<td>-0.076</td>
</tr>
</tbody>
</table>

* = Standardized

<table>
<thead>
<tr>
<th>Random part components</th>
<th>Generating estimate</th>
<th>Formulating estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>(S^2 \beta_0)</td>
<td>0.430</td>
<td>0.086</td>
</tr>
<tr>
<td>(S^2 \beta_1)</td>
<td>0.002</td>
<td>0.003</td>
</tr>
<tr>
<td>(S^2) Text quality*</td>
<td>0.183</td>
<td>-0.003</td>
</tr>
</tbody>
</table>

* = Standardized

---

\(^7\) Remember: \(\text{logit}(P) = \log(P/(1-P))\)

\(^8\) 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 tasks

<table>
<thead>
<tr>
<th>Random part components</th>
<th>Reading the assignment estimate</th>
<th>Planning estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$S^2 \beta_1$</td>
<td>$S^2 \beta_2$</td>
</tr>
<tr>
<td>$S^2 \beta_0^*$</td>
<td>1.023</td>
<td></td>
</tr>
<tr>
<td>$S^2 \beta_1$</td>
<td>0.226</td>
<td>0.051</td>
</tr>
<tr>
<td>$S^2 \text{Text quality}^*$</td>
<td>0.015</td>
<td>0.010</td>
</tr>
</tbody>
</table>

* = Standardized

<table>
<thead>
<tr>
<th>Random part components</th>
<th>Generating estimate</th>
<th>Formulating estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$S^2 \beta_1$</td>
<td>$S^2 \beta_2$</td>
</tr>
<tr>
<td>$S^2 \beta_0^*$</td>
<td>0.147</td>
<td></td>
</tr>
<tr>
<td>$S^2 \beta_1$</td>
<td>0.003</td>
<td>0.004</td>
</tr>
<tr>
<td>$S^2 \text{Text quality}^*$</td>
<td>0.002</td>
<td>0.003</td>
</tr>
</tbody>
</table>

* = Standardized