Relating self reports of writing behaviour and online task execution using a temporal model

**Marion Tillema · Huub van den Bergh ·
Gert Rijlaarsdam · Ted Sanders**

Abstract Current theory about writing states that the quality of (meta)cognitive processing (i.e. planning, text production, revising, et cetera) is, at least partly, determined by the temporal distribution of (meta)cognitive activities across task execution. Put simply, the quality of task execution is determined more by *when* activities are applied than by *how often* they are applied. *Planning* and *revising* are two extreme writing styles, in which (meta)cognitive activities are temporally differently distributed across the writing process. Planners are writers who generate plans before text production. Revisers use text production as a means to arrive at a content plan. The present study investigates the question whether the online (meta)cognitive processing of secondary school students during writing tasks, as measured by think aloud techniques and keystroke logging, can be predicted by their responses to an offline questionnaire which measures to what degree students considered themselves to be planners and revisers. It was expected that different reported writing styles would entail different temporal distributions of six (meta)cognitive activities: reading the assignment, planning, text production, reading own text, evaluating own text and revising. This hypothesis was partly confirmed. The results show that the online temporal distributions of reading the assignment and planning are different for different degrees of reported writing styles. On the basis of these results, the validity of both the questionnaire and the concept of planner and reviser styles are discussed.

Writing a coherent and readable text involves handling many (meta)cognitive activities, such as generating, planning, translating ideas into language and making revisions when

M. Tillema (✉) · H. van den Bergh · T. Sanders
Utrecht Institute of Linguistics OTS, Utrecht University, Trans 6, 3512 JK Utrecht, The Netherlands
e-mail: M.Tillema@uu.nl

H. van den Bergh · G. Rijlaarsdam
Research Institute for Child Development and Education, University of Amsterdam, Spinozastraat 55, 1018 HJ Amsterdam, The Netherlands

needed, while keeping rhetorical goals in mind and staying aware of the intended audience (Flower and Hayes 1980; Hayes and Flower 1980; Hayes 1996). The writing process is not a linear chain of actions in which planning, generating, text production and revising, for example, are carried out in consecutive phases. Rather, it is a recursive process, in which cognitive activities may be re-applied during any phase of the writing process (Hayes 1996). Sometimes a fully developed content plan is subsequently translated into text; in other cases, development of the text plan correlates with the development of the written-down-text. In the latter cases, writing is an act of discovering what to say (Galbraith 1996; Hayes 1996). In short, there are numerous possible configurations of (meta)cognitive activities (Rijlaarsdam and Van den Bergh 1996; Van den Bergh and Rijlaarsdam 2001; Van Weijen et al. 2008).

Torrance et al. (1994) present two extreme writing styles, with different configurations of writing activities. They identified *planners*, "who planned extensively and then made few revisions" on the one hand, and *revisers*, "who developed content and structure through extensive revision" on the other hand. In addition, they identified so-called *mixed strategy writers*, who applied both planning and revising activities extensively. Similar strategies are found in Torrance et al. (1999, 2000), although they are labeled differently, together with a number of additional strategy types. Biggs et al. (1999) present a typification of writing strategies similar to the planner/reviser distinction. On the one hand, *engineers* plan extensively before commencing with text production. On the other hand, *sculptors* start text production in a relatively early stage of the writing process, without much planning preceding it. The content plan develops as the text develops. The produced text is subsequently revised until it fits what the writer wants to say. Kieft et al. (2006, 2008), finally, also use these two writing styles in their research. They quote Galbraith and Torrance (2004) to describe the *planning strategy* as a strategy "in which writers concentrate on working out what they want to say before setting pen to paper, and only start to produce full text once they have worked out what they want to say" and the *revising strategy* as a strategy "in which writers work out what they want to say in the course of writing and content evolves over a series of drafts" (Kieft et al. 2008, p.380).

Planners and revisers, then, by definition have different configurations of planning activities, text production activities and revision activities. Other cognitive activities might be expected to have different distributions, too. Reading the assignment, for example, probably occurs more early on in the writing process for extreme planners, but in later stages of task execution for typical revisers. After all, typical planners will think about whether the text matches the assignment during planning stages, while typical revisers will think about this once a text has been produced. The moments of occurrence of all of these activities is, of course, interrelated. If planning happens early on in the writing process, for example, revising cannot.

In short, planners and revisers apply the various (meta)cognitive activities at different moments during the writing process. This leads us to observations made by Rijlaarsdam and Van den Bergh (1996) and Van den Bergh and Rijlaarsdam (1996). They demonstrated that the occurrence of (meta)cognitive activities varies across task execution. Structuring activities (a subcomponent of planning), for example, are on average more likely to occur a short while after the start and also towards the end of task execution, but less likely to occur during middle stages of the writing process. They also showed that the distributions of cognitive activities differ between individual writers. Some writers, for example, follow the average distribution of structuring activities, while others tend towards a different distribution. One, for example, in which structuring activities are hardly used at the start of the task, a little more during middle stages, and mostly during the final phases of the

writing process. Finally, Rijlaarsdam and Van den Bergh (1996) and Van den Bergh and Rijlaarsdam (1996) demonstrated that the relation between cognitive activities and text quality varies across task execution. Structuring activities, for instance, were shown to be more effective when they occurred during early stages of task execution (i.e. the correlation between structuring and text quality is at its highest at the start of the writing process) and less effective when they occurred towards the end of task execution. They therefore advocated a temporal analysis of activities over the writing process: analyses of cognitive processing during writing should take the moment(s) at which cognitive activities occur into account. The validity of this temporal approach was confirmed by Breetvelt et al. (1994), Van den Bergh and Rijlaarsdam (2001), Van den Bergh et al. (2009) and Van Weijen et al. (2008), who also demonstrated that differences between writers in terms of distributions of cognitive activities explain differences in the quality of the texts produced. The temporal approach of writing processes has become a dominant view in writing research (Leijten and Van Waes 2006; Olive et al. 2008).

The configuration and temporal distribution of (meta)cognitive activities is an online characteristic: we can establish it by measuring what happens during the process of task execution. A common method for measuring the process of task execution (during writing, but also during other tasks, for example reading tasks or mathematic problems) is the use of think aloud techniques (cf. Cromley and Azevedo 2006; Rijlaarsdam and Van den Bergh 1996; Roca de Larios et al. 2008; Van den Bergh and Rijlaarsdam 1996; Van Weijen et al. 2008; Veenman and Spaans 2005). Another method for concurrent measurements of writing processes is keystroke logging (e.g. Leijten and Van Waes 2006; Strömqvist et al. 2006). Online measurements have been shown to have predictive value for the quality of the output of task execution (Cromley and Azevedo 2006; Torrance et al. 1999; Van der Stel and Veenman 2008; Van Weijen et al. 2008; Veenman et al. 2003). This output may be text quality (for writing), but also test scores (in other domains, such as reading and mathematics).

However, there are also numerous studies where writing behaviour is measured independently from the writing process. Questionnaires about different aspects and/or configurations of writing processes (Kieft et al. 2006, 2008; Lavelle et al. 2002; Torrance et al. 1994, 1999, 2000) are an example of such offline measures. The use of offline measurements have been criticised as inaccurate reflections of the underlying process. Russo et al. (1989), for example, found the contents of retrospective protocols to be incomplete and partly fabricated. Their opinion is shared by Veenman et al. (2003) and Cromley and Azevedo (2006). In both studies, offline reports were related to online data, the latter in the form of total or relative frequencies of strategy-related verbalisations in concurrent data (Cromley and Azevedo 2006; Veenman et al. 2003) or by proportions of indicated strategy use in a concurrent multiple-choice tool (Cromley and Azevedo 2006). They found that relations between offline reports and online task execution were weak or absent.

However, analyzing online metacognition by establishing frequencies of metacognitive verbalisations runs counter to the idea that the quality of online task execution is determined by the *temporal distribution* of (meta)cognition across the writing process. Possibly, this could form an explanation for the absence of (substantial) relations between offline and online data in these studies.

Torrance et al. (1999) indeed showed a correspondence between questionnaire outcomes and online data, the latter being analyzed in terms of distributions. Participants in their study completed a questionnaire about their writing behaviour. On the basis of this questionnaire, participants were categorised into one out of three possible strategy groups. The questionnaire outcomes in this study predicted online writing behaviour. This online behaviour was analyzed in terms of distributions of (meta)cognitive activities, such as planning, translating

ideas into language and revising. Torrance et al.'s (1999) study, then, suggests that offline reports of writing behaviour[1] can, at least to some extent, be used as predictors of a general tendency towards a particular online distribution or configuration of cognitive activities.

A Writing Style Questionnaire developed by Kieft et al. (2006, 2008) measures reported degrees of planner- or reviser-type writing behaviour within individuals. Contrary to Torrance et al. (1999), Kieft et al. (2006, 2008) do not categorize writers, i.e. writers are not either planners or revisers. Rather, the two dimensions (i.e. the planner and reviser dimension) are seen as scales, on both of which the degree to which it applies to an individual writer is expressed. Kieft et al. (2008) provide some evidence which seems to suggest a degree of validity for this questionnaire. They tested students' writing style by means of the Writing Style Questionnaire. Subsequently, all students participated in a lesson series on writing. One group of students (consisting of both students for whom planning was the dominant writing style and student for whom revising was the dominant writing style) received instruction that matched a planning style and another group of students (again consisting of both students for whom planning was the dominant writing style and student for whom revising was the dominant writing style) received instruction that matched a revising style. They found that study outcomes (i.e. the quality of the texts which the students wrote) increase if writing lessons match the most dominant writing style in students' responses to the questionnaire. This is, however, indirect evidence for the assumption that the Writing Style Questionnaire is a predictor of online writing behaviour. There has, to date, been no research to test whether higher or lower degrees of reported planner- or reviser-type behaviour do indeed predict different online configurations and temporal distributions of (meta)cognitive activities. This is investigated in the present study. It may be assumed, for example, that 'high planners' (according to the Writing Style Questionnaire) perform more planning activities at the start of task execution than 'low planners'. After all, we know that planning activities are most effective during initial stages of the writing process (Van den Bergh and Rijlaarsdam 2001). Similarly, it may be assumed that 'high planners' will apply less planning activities than 'low planners' at stages in the writing process during which planning activities are less effective: towards the end of task execution. 'High revisers', on the other hand, will generally apply more planning activities at the end of task execution than 'low revisers'. After all, typical revisers use text production to arrive at a plan of what to say. As a consequence, 'high revisers' will also apply more revision activities at the end of task execution than 'low' revisers. In the same vein, we predict that different scores on the planning and revising dimension in the Writing Style Questionnaire (Kieft et al. 2006, 2008) are related to different distributions (across task execution) of the other (meta)cognitive activities which occur during writing.

Method

Participants

The participants were fourteen- and fifteen-year-old students ($N=20$; 10 female and 10 male). They were from three different third-year-forms at the same school for pre-university

---

[1] Offline reports of writing behaviour are essentially reports of what learners know about their writing activities and strategies. They are, in other words, measures of what Flavell (1979) calls *metacognitive knowledge about strategy,* or what Zohar and David (2009) call *Meta-strategic Knowledge*: "an awareness of the type of thinking strategies being used in specific instances".

secondary education. They were recruited by means of a call for volunteers, which was distributed by their Dutch language teacher. All participants were native speakers of Dutch. They received a small financial compensation for their participation. Parental consent was obtained.

Tools and procedures

The students completed four writing tasks. In addition, they completed an offline questionnaire to measure reported writing behaviour. They performed all tasks individually in a university room, in the presence of a test leader.

Writing tasks

All students wrote four argumentative essays in Dutch, their mother language, on topics such as 'camera surveillance in inner city areas' or 'legalisation of soft drugs'. They completed all their essays during one session, with a short break of about 15 min between assignments. The sequence of topics was systematically balanced across participants.

The assignments consisted of a brief statement of topic, audience (peers), medium (the school paper) and purpose (to convince the readers of your point of view), followed by a series of quotes (factual information as well as opinions) that were related to the topic, of which two had to be used in the essay. All assignments were tested with third year students of pre-university secondary education during a pilot study in 2005. They were also successfully used by Van Weijen et al. (2008, 2009). The essays had to be about half a page in length (which is about 250–300 words). An example of an assignment can be found in Appendix A.

The available time for each essay was 30 min. The mean writing time was 20.13 min (SD=5.89, $Min.$=7.80, $Max.$=32.15). The time spent on each task was related to the order in which tasks were completed (χ^2=7.23, df=1, p<0.01). The mean writing time for the first essay in the session was 21.74 min, while the mean writing time for the last essay was 18.93 min. That students spent less time on the last task in the session than on the first task can probably, for a large part, be explained by the fact that students generally needed less time for reading the assignment during later tasks: a fairly large portion of the instruction text (e.g. description of audience, medium and purpose) was identical in all tasks.

The students wrote the essays on a computer using Microsoft Word. They had to think aloud during the process of task execution. All writing sessions were video-taped. The writing sessions were also recorded by means of keystroke logging (Inputlog: Leijten and Van Waes 2006), in order to obtain more detailed information on text production and revision activities.

Writing style questionnaire

The students also completed Kieft et al.'s (2006, 2008) Writing Style Questionnaire. This questionnaire measures reported degrees of planning and revising style. It is specific to the domain of argumentative writing, in that participants are asked how they would handle writing an argumentative essay about the tobacco industry. This 'tobacco' task, which was not actually carried out by the participants, is very similar to the four writing tasks performed by the students in the present study in terms of text type and intended medium.

The questionnaire consisted of thirty-six statements about writing strategy. Thirteen of these items reported planning-type behaviour and twelve of these items reported revising-type behaviour. The remaining eleven items are fillers. Students had to indicate in how far each statement pertained to them, by checking a box on a five-point scale. On the basis of

Planning

Before I start writing, I want to have it clear which information to put in the text. Therefore, planning is important to me.

If I have to write a text, I spend a lot of time on thinking about my approach.

I always make a text schema before I start writing.

If I have to write something, I jot down some notes, which I work out later.

Before I start writing a text, I write something on a scribbling pad, to find out my opinion about the topic.

* Planning is of no use to me.

* When I start writing, I don't yet have a clear idea of what will be in the text.

Before I start writing, I have a clear picture of what I want to achieve with the readers.

I need to have my thoughts clear before I am able to start writing.

Before I write a sentence down, I already have it in my head.

* When I am writing, I sometimes write down pieces of text of which I know that they are not completely right yet. Still, I prefer to go on writing at that point.

* When I read over my texts, I usually find a lot to improve.

* When I read over my texts, they are sometimes very chaotic.

Revising

* I always start writing straight away: I don't need to know exactly what I will write or how the text will be built-up. That will become clear as I write.

When my text is ready, I elaborately read through it and make improvements: a lot can still be changed at that point.

Fig. 1 Items in the Writing Style Questionnaire (Kieft et al. 2006, 2008), sorted according to which dimension they measure

During writing I regularly check if my text does not contain any sentences which are incorrect or too long.

While writing my text, I continually ask myself if readers will be able to follow it.

For me, writing is a way to get my thoughts clear.

* I usually hand in my text without checking if its organization is in order.

If I read over my texts, and rewrite my texts, it occurs regularly that I drastically change their organization.

Before I hand in a text, I always check if its build-up is logical.

* I never pay much attention to whether I have forgotten to put any sentences or ideas in a text.

When I rewrite a text, the content usually changes drastically, too.

When I finish a text, I usually need to read through it carefully, to check if there is no superfluous information in it.

I never pay much attention to whether I am satisfied with my texts.

Fillers

I write and rewrite my text sentence per sentence. Only if I am completely satisfied with a sentence, do I proceed with writing.

When I am writing, I find it hard to organize my thoughts.

Only if my text is complete, do I read what I have written.

If finally I have an approximate idea of what to say in my text, the words will flow out of my pen.

When I write, I stop writing after every few sentences to read what I have just written.

Fig. 1 (continued)

| I try to write a correct version of my text in one go, so that I hardly have to make any alterations when it's finished. |
| When I write a text, I find it hard to come up with ideas. |
| When I am writing, I often find that all kinds of new ideas pop into my head. |
| For writing tasks, I do not find it very hard to think of arguments to support my point of view. |
| The texts which I write are usually not very original. |
| I make sure that every sentence is perfect, before I start with the next sentence. |
| When my text is finished, the only thing I do is check for language or spelling mistakes. |

Fig. 1 (continued)

their questionnaire responses, participants received scores for both the planning dimension and the revising dimension. They could therefore score equally high or low on both dimensions, or one of the two dimensions could be dominant. Figure 1 features all questionnaire items, which are sorted according to the dimension they pertain to. In the actual questionnaire, the items were presented in random order. For the present study, the questionnaire was in Dutch, the students' mother language.

Analyses

All think aloud data were transcribed and segmented, which were completed by the Inputlog recordings. A new segment in the protocols reflected a switch to a different (meta)cognitive activity within a participant's writing session (Van den Bergh and Rijlaarsdam 2001).

All segments were coded according to a coding scheme (adapted from Breetvelt et al. 1994). One think aloud protocol (number of segments=518) was coded by two researchers. The intercoder agreement (Kappa) was 0.85. The coding scheme (see Table 1) consists of fourteen categories. Five of these categories reflect planning activities, namely monitoring, goal setting, generating content, structuring (which involves the selection and evaluation of propositions which have been generated but not (yet) translated into text) and metacomments. One category involved subcodings, namely 'Revising'. Revisions were subcoded as 'automated corrections' when they involved corrections of typographic errors. They were typically errors which are made as a result of the use of a keyboard, which seem to be corrected almost automatically. An example would be: a writer types 'almots', and immediately corrects this error by giving to backspaces and typing 'st', so that it now reads 'almost'. All revisions which were not 'automated corrections' were subcoded as 'conceptual revisions': they involve alterations which are made by the writer in a non-automated way. That is, they involve actual alterations at the level of spelling or content. In the remainder of this article, we will mean conceptual revisions if we use the term 'revision' or 'revising'.

Table 1 Coding categories in the coding scheme

	Coding category	Description	Example
Reading the assignment	Reading the instruction text and documentation	Reading (part of) the task instructions	"Write an essay in which you…"
Planning	Monitoring	Verbalizations indicating a steering capacity which governs the writing process. Mostly self-instructions	"I'm going to read what I've written so far."
	Goal setting	The formulation of goals which the text has to satisfy	"The text should be convincing."
	Generating	Generating ideas for content or form	"Something about the disadvantages of camera surveillance…"
	Structuring	Evaluating and arranging ideas	"Something about adults? No, that's not relevant."
	Metacomments	Evaluations of a student's own writing process	"I should have made an outline of the text before I started."
Text production	Text production	Production of new text	"Camera surveillance invades people's privacy." (Verbalisations usually occur parallel to the activity of typing.)
Reading own text	Reading produced text	Reading (part of) the produced text, at any given moment during the writing process	"Surveillance cameras do not increase public security."
Evaluating own text	Evaluating produced text	Evaluation of produced text	"The largest part is about backdraws."
Revision	Revising	Making changes to the text produced so far. Changes can be made at word, sentence or text level.	Moving a set of sentences from the body of the text to the introduction.
Other	Pauses	Silence or interjection	"eeeerrr"
	Interaction with test leader	Interaction between test taker and test leader	"Could I open a window?"
	Physical activity	Physical activity	Taking a sip of tea.
	Navigation	Moving through the document: arrow buttons or mouse movements	Moving the cursor some lines back.

As different degrees of reported planner- or reviser-style behaviour entail different configurations of the complete writing process, we expect different distributions for all of the activities listed in Table 1: reading the assignment, planning, text production, reading own text, evaluating own text, and revision. An exception is formed by the activities categorized as 'OTHER'. For these activities, there is no conceptual link with planner and reviser styles. The same applies for the subcategory of 'automated corrections'. These five activities (pauses, interactions with test leader, physical activity, navigation and automated corrections) were therefore not included in the analysis.

Figure 2 shows an example of (a part of) a protocol and illustrates how the think aloud data en the Inputlog data were integrated. All information in the column labelled 'Typing' are derived from Inputlog. This column contains 'text production' and 'revising' activities.

Segment number	Reading	Verbalizations	Typing	Pausing	Other	Code
17		even een titel erboven (*I'll insert a title*)				Monitor
18		having	Having childere			Text production
19		chil	[BS 1] [BS 1] [BS 1]			revision (automated correction)
20		children yes or no	ren, yes or no?			text production
21	ROT	having children yes or no				Reading own text

Fig. 2 Part of a completed protocol

The protocols consisted of seven columns. The column labeled 'Reading' was used for indicating if any reading activities (reading the assignment—RA—or reading own text—ROT) were taking place. The 'Verbalizations' column contained everything which was said out loud by the student, except interjections, such as "uhm". The 'Typing' column contained all text production as registered by Inputlog. There were three categories in this column, namely the production of new text, revisions (indicated by Inputlog as [BS] for backspace or [DEL] for if the delete button was pressed), and navigation (by means of mouse movements or arrow buttons). The 'Pausing' column contained all silences and interjections. The column labeled 'Other' mostly contained descriptions of physical activities, for example 'takes a sip of his drink'. One row is one protocol segment. As such, this transcription method allows for parallel actions. Text production and verbalizations, for example, often occur simultaneously. The 'Typing' column would contain the text production as registered by Inputlog. The codings in the last column were in reality

numbers. Code 01, would stand for 'reading the assignment', for example, and code 02 would stand for monitoring. English translations of Dutch verbalizations are given in italics.

Due to technical deficits, there were less than four writing sessions available for analysis for three participants. For one participant, two writing sessions were included in the analysis. For two other participants, three writing sessions were included. For the remaining seventeen participants, all four writing sessions were available.

Modelling (meta)cognitive activities across task execution

The first step in the analysis is to model the (online) occurrence of (meta)cognitive activities temporally, that is, as a function of the moment in the writing process. We constructed this time variable by splitting each protocol into five equally long episodes in terms of numbers of segments (cf. Roca de Larios et al. 2008; Van den Bergh and Rijlaarsdam 1996). A protocol of 330 segments, for example, would be analyzed as five episodes consisting of 66 segments each. By using episodes we achieved standardisation: it allowed us to compare different writing processes between (and within) individuals in terms of start, middle and end of task execution. After all, episode 3, for example, reflects the middle part of the writing process for each protocol, no matter if it contains segments 133–199 in a protocol of 330 segments, or segments 101–150 in a protocol of 250 segments. The (meta)cognitive activities which are the dependent variables in our analysis (i.e. reading the assignment, planning, text production, reading own text, evaluating own text, and revision) were all expressed as proportions of the total number of segments for each episode. For instance, if an episode consisted of 80 segments (which would mean that the entire writing process consisted of 400 segments), and 10 of these segments were coded as 'reading the assignment', then the proportion for reading the assignment in that episode would be 0.25.

A multilevel regression model has been applied to model the occurrence of the (meta) cognitive activites at each episode, as episodes are nested within writers (Van den Bergh et al. 2009). The analysis was conducted with MLwiN software for multilevel models. In effect, a longitudinal model is in operation, as it concerns changes in occurrence during the writing process: proportions of the applied (meta)cognitive activity may be different during each new episode. Therefore, the occurrence of each of the six online activities (A) had to be described as a function of episode, i.e. $A = f(episode)$. Note, however, that this function f does not need to be identical for all individuals i: $A = f_i(episode)$.

This function, f, can take many forms (Goldstein 1979; Healy 1989). For this study, polynomial models were preferred because of their flexibility. Depending on the number of coefficients (and their numerical values), polynomials can take almost any shape. As such, they can be used to model various kinds of growth patterns.

Growth across task execution is not necessarily linear. For instance, text production activities may occur relatively little during first and last episodes of the writing process, but a lot during the middle part of task execution (e.g. during episode 3) Therefore, non-linear terms (e.g. quadratic or cubic terms) can also be included in the model: the occurrence of an activity (at each episode) is described as powers of episode ($episode^0$, $episode^1$, $episode^2$, …). The number of parameters needed to describe the observed activities (in each episode) is considered an empirical matter. That is, a next power of episode is only included in the model if it has a significant contribution in the description of an activity *and* if all lower powers are significant as well (see, Van den Bergh and Rijlaarsdam 1996). For example, 'episode' to the second power can only be added to the model, if the linear term ($episode^1$) has reached significance.

In order to meet the requirement that the function f is allowed to differ between individuals, not only are the regression coefficients of the powers of episode estimated, but

also the variance of these parameters. That is, the variance of the intercept (writers differ in occurrence at episode = 0), the variance of the linear component (writers differ in linear change over the writing process), et cetera.

These variance components are in fact the variances of residuals which characterize the occurrence of activities of a specific writer. Therefore, the differences between individuals can be explained by individual characteristics like their offline planner and reviser scores. Adding these offline scores, then, is the final step in the construction of a multilevel regression model, in which episode and the individual planner or reviser scores were the explanatory variables. Of course, the effect of these offline scores is not (necessarily) constant across task execution. (For instance: we expect differences in process execution due to higher planner scores to be larger at the start of task execution than during later parts.) Therefore, interaction effects between the offline scores and the time variable (episode) on the dependent variable were also calculated. The complete multilevel regression model, as construed in MLwiN and as used for explaining the occurrence of each of the six (meta)cognitive activities, can be found in Appendix B.

Results

Internal consistency was calculated for the Writing Style Questionnaire. Cronbach's alpha is .72 for the items on the planner dimension and .64 for the items on the reviser dimension. These reliabilities, which are similar to the reliabilities found by Kieft et al. (2006, 2008), justify aggregating the items for each dimension to calculate mean scores per dimension per student. As the two dimensions are only moderately correlated ($r=.39$), planning- and revising-type behaviour can be identified separately in the Writing Style Questionnaire data (see also Kieft et al. 2006, 2008).

Table 2 features descriptive information about the data. It shows that writing processes are on average 346.68 segments long, but that there is great variation ($SD=137.88$). In addition, Table 2 shows that text production, which on average takes up about 131 segments (which is about 38% of the segments), occurs far more often than the other activities in the analysis, which on average occur in 0,5% to 5% of the segments.[2] However, the data show a relatively larger range for the more infrequent activities than for text production. The standard deviations for 'reading the assignment', 'planning', 'reading own text', 'evaluating own text' and 'revising' are in most cases larger than their means. For these activities then, there seems to be a lot of variation due to episode (i.e. moment during the writing process) and student. This variation is, of course, to be explained by the results of the regression analyses.

The results of the regression analyses show that for the six online activities which were analyzed, the average occurrence indeed varies across task execution, i.e. due to 'episode'. (See Appendix C for parameter estimates.) The proportion with which reading the assignment, planning, text production, reading own text, evaluating own text and revising are applied, is significantly different if episode is the explanatory variable.

Pseudo R^2 was calculated for the six models which were constructed to explain the occurrence of online activities with the 'episode' variables. The outcome is presented in

---

[2] 'Pausing' and 'automated corrections' were relatively frequently occurring activities. This explains why the accumulated proportions for the six (meta)cognitive activities presented in Table 2 do not approximate 1. (.0569 + .0544 + .3791 + .0295 + .0061 + .0005=0.497)

Table 2 Descriptive statistics for the Writing Questionnaire scores, and (meta)cognitive activies in the concurrent protocols (absolute numbers and proportions)

Descriptives

	M	SD	Min.	Max.
Score *planning* dimension (Writing Style Questionnaire)	2.45	.66	1.50	4.13
Score *revising* dimension (Writing Style Questionnaire)	3.16	.57	2.25	4.38
Number of segments per participant, per task	346.68	137.88	134	767
Number of segments containing 'reading the assignment' activities per participant, per task	18.96	14.48	5	75
Number of segments containing 'planning' activities per participant, per task	19.21	19.71	0	96
Number of segments containing 'text production' activities per participant, per task	131.16	56.77	44	282
Number of segments containing 'reading own text' activities per participant, per task	11.18	11.65	0	54
Number of segments containing 'evaluating own text' activities per participant, per task	2.39	3.21	0	13
Number of segments containing 'revising' activities per participant, per task	0.20	0.65	0	4
Proportion of 'reading the assignment' per episode	.0569	.05793	.00	.31
Proportion of 'planning' per episode	.0544	.06143	.00	.40
Proportion of 'text production' per episode	.3791	.09192	.10	.58
Proportion of 'reading own text' per episode	.0295	.04004	.00	.25
Proportion of 'evaluating own text' per episode	.0061	.01636	.00	.15
Proportion of 'revising' per episode	.0005	.00367	.00	.04

Means (*M*), standard deviations (*SD*), minimum and maximum scores/numbers of verbalisations (*Min.; Max.*)

Table 3. Only for one of the activities, R^2 is low (reading own text: $R^2=0.18$). As this is not a crucial activity in our analysis, this is a relatively minor problem. For the other five activities (reading the assignment, planning, text production, evaluating own text and revising), R^2 proved to be satisfactory.

Figure 3 shows the average distributions of all six activities. The relation between online activities and episode was analyzed in logits. As logits are hard to interpret, we transformed them into proportions, in order to interpret the results as probabilities of occurrence. In the figures below, then, we present distributions of activities in proportions.

Table 3 R^2 describes the fit of the constructed regression models

| Activity	R^2
Reading the assignment	.69
Planning	.59
Text production	.58
Reading own text	.18
Evaluating	.78
Revising	.81
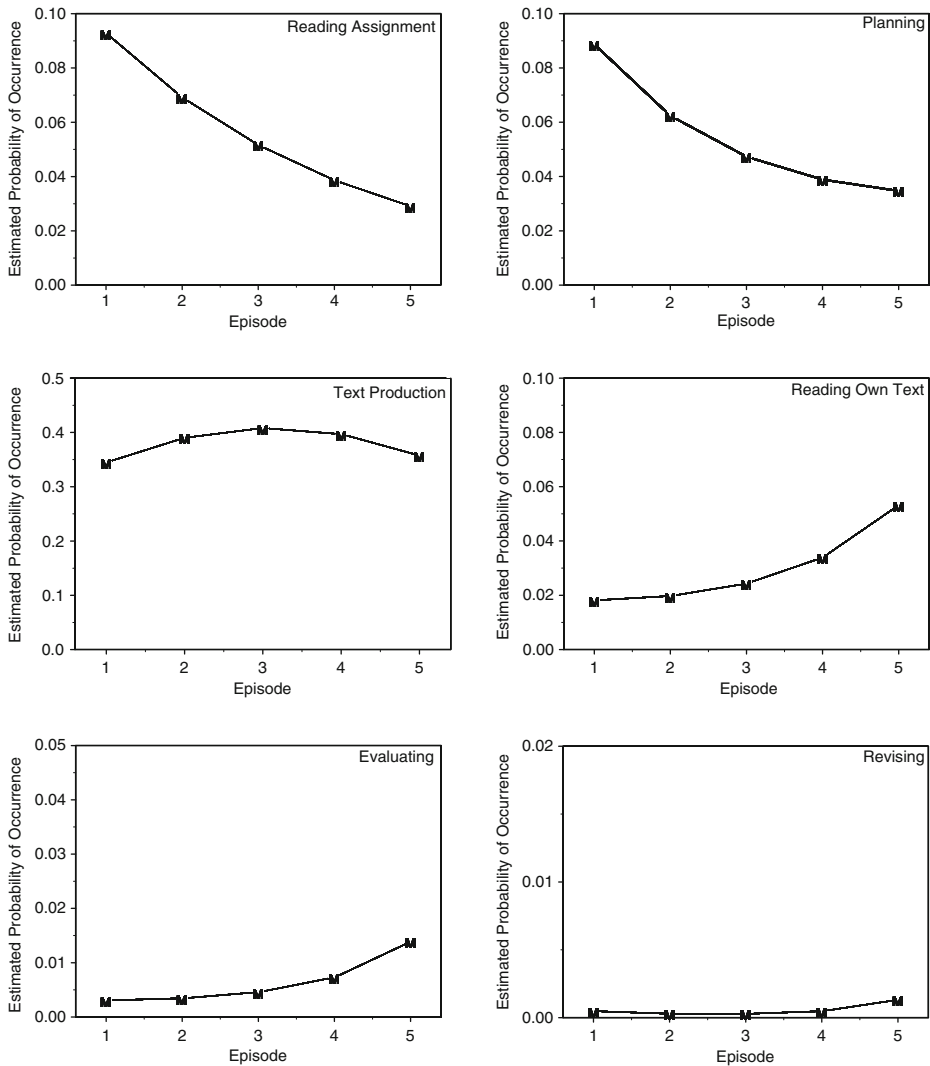
Fig. 3 Average distributions across the writing process for six activities: Reading the Assignment; Planning; Text Production; Reading Own Text; Evaluating Own Text; Revision

Figure 3 illustrates that the probability that 'reading the assignment' occurs is highest in episode 1, that is, at the start of task execution. Thereafter, its probability of occurrence declines (with slightly different amounts) with every next episode. To put it more simply: reading the assignment happens most often at the start of the writing process (episode 1), and least often at the end of task execution (episode 5). The same pattern applies for planning activities. Text production activities are distributed differently across task execution. They are already quite likely to occur at the start of task execution, with an estimated probability of almost .35. After the start of task execution, the probability increases, reaching its peak at the middle stage of the writing process. After episode 3, there

is a slight decrease towards the end of task execution. Text production activities, in short, occur quite frequently across the entire writing process, but are most frequent during the middle part of task execution. Both reading and evaluating own text occur least at the start and most at the end of task execution, although they are still very unlikely to occur there. This is even more the case for revision activities. Although they are very slightly more probable during episode 5, revision activities on average occur infrequently at any moment during writing.

The next question now is if these distributions vary according to offline planner and reviser scores. Table 4 gives an overview of whether significant effects were found. The second and third column show if there is a significant main effect of planner or reviser scores on the number of activities applied. Such an effect would mean that differences in offline scores are related to differences in the number of times that an activity occurs during the entire writing process. The fourth and fifth column show if there are significant interaction effects of planner/reviser and episode. The existence of such an effect would mean that the effect of higher or lower offline scores varies across episodes. It would mean, in other words, that distributions are different for different planner or reviser scores. (See Appendix D for parameter estimates.)
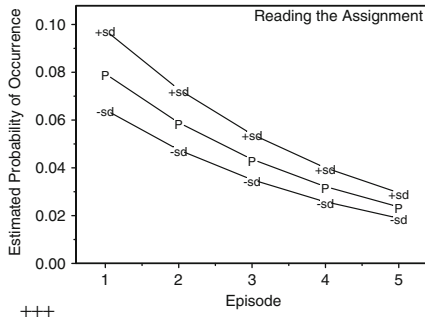
Table 4 shows that there is a main effect of planner scores on four of the six online activities, namely reading the assignment, planning, text production and revising. Significant main effects of reviser scores exist for the activities planning and reading own text. Interaction effects could be established in two cases: 1) for online planning activities, the effect of planner scores is different for various episodes, and 2) for reading the assignment, the effect of reviser scores is different for various episodes. The direction of the effects can be inferred from the regressions weights, and are illustrated in Fig. 4. Again, the logits were transformed into proportions to facilitate the interpretation of the graphs. Figure 4 shows the variations in the distributions of (meta)cognitive activities according to variations in offline planner and reviser scores. Each graph contains three lines: one (P or R) reflecting the distribution of the specific activity for students with average planner or reviser scores, one (+sd) reflecting the distribution for students with a planner or reviser score of one standard deviation above the average score, and one (−sd) reflecting the distribution for students with a planner or reviser score of one standard deviation below the average score. No effects, in any form, were found on the activity 'evaluating own text'. This activity is therefore not featured in Fig. 4. For text production, reading own text and

Table 4 Overview of the effects of offline planner and reviser scores and episode on six (meta)cognitive activities measured online
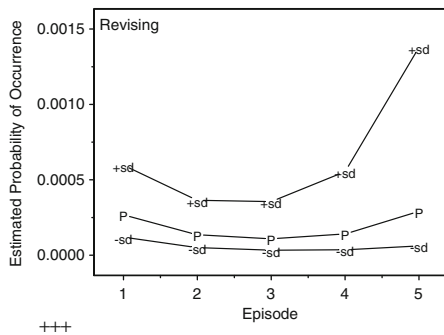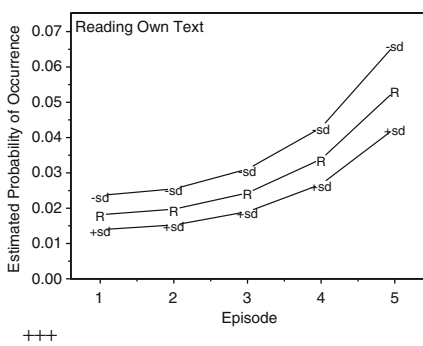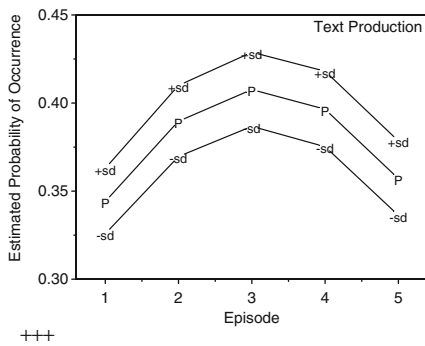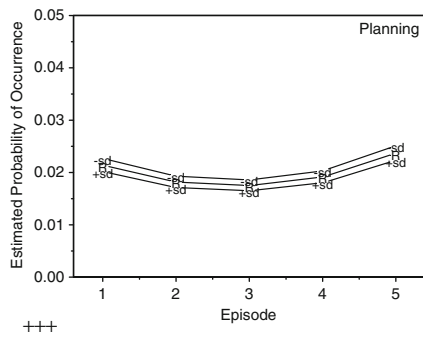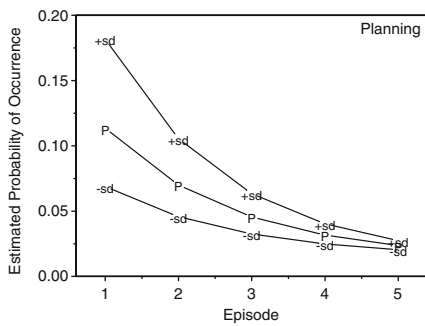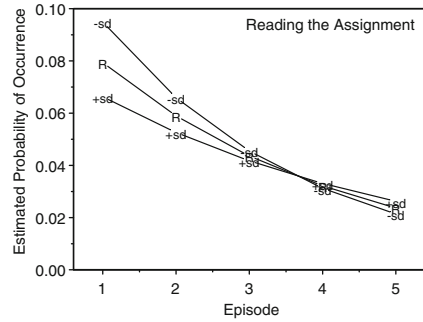
	Effects of WSQ scores and episode			
	Planner	Reviser	Planner*episode	Reviser*episode
Reading the assignment	V	–	–	V
Planning	V	V	V	–
Text production	V	–	–	–
Reading own text	–	V	–	–
Evaluating own text	–	–	–	–
Revising	V	–	–	–

V: a significant effect was observed. –: no significant effect was observed

Variation in distributions due to different *planner* scores.

Variation in distributions due to different *reviser* scores.

+++

+++

+++

+++

+++

+++

revising, we observed an effect of either only the planner or only the reviser scores. Therefore, only one graph is displayed for each of these three activities. The graphs marked with + + + illustrate the variations due to episode and planner/reviser scores as observed, although only the main effects (i.e. effects of only the planner or reviser score) were significant.

Students with a higher planner (+sd) score generally read the assignment on fewer occasions than average. Students with a lower planner (−sd) score read the assignment more frequently than average. Although the difference between high and low planners seems to become smaller as the writing process progresses, there was no significant interaction effect of planner score and episode. In other words, no evidence could be found that the effect of a higher or lower planner score is different for different episodes. We have to assume that it is stable across task execution. The effect of reviser scores on the occurrence of reading the assignment does vary over time, i.e. there is a significant interaction between the variable 'reviser score' and the variable 'episode'. At the start of task execution, students with lower reviser scores (−sd) are (slightly) more likely to read the assignment than students with higher reviser (+sd) scores. From episode 2 onwards, however, this effect is reversed. From that moment on, students with lower reviser scores are less likely to read the assignment than students with higher reviser scores. The difference (in terms of the probability that 'reading the assignment' occurs) between students with higher and lower reviser scores grows larger towards the end of task execution.

For planning activities, the effect of higher or lower planner scores varies across task execution. Students with higher planner scores (+sd) are more likely to apply planning activities at the start of task execution than students with lower planner scores. This changes fairly soon after the start of task execution, so that towards the end of task execution, students with higher planner scores are less likely to apply planning activities than students with lower planner scores. The effect or higher or lower reviser scores was not found to vary across task execution for planning activities. An overall effect was established: the higher the reviser score, the more planning occurred in total during the writing process.

For text production, there was a significant main effect of planner scores: the higher the offline planner score, the more text production activities occur. For reading own text, there was a significant main effect of reviser scores: the higher the offline reviser score, the less 'reading own text' occurs. For revision activities, the difference between high and low planners seems larger at the start and the end of task execution. This interaction effect was, however, not significant. There was a significant main effect of planner scores on revising: the higher the offline planner scores, the more revising activities occur.

Discussion

We predicted that differences in reported planner and reviser styles as measured by Kieft et al.'s (2006, 2008) Writing Style Questionnaire were related to different distributions of various (meta)cognitive activities over the course of the writing process. Results indicate that the occurrence of six (meta)cognitive activities—reading the

assignment, planning, text production, reading own text, evaluating own text and revising —varies across task execution. Activities are more likely to occur during some episodes than during others. Different distributions due to reported writing style were found for two out of the six activities which were analyzed, namely for reading the assignment and planning. For three other activities, namely text production, reading the assignment and revising, the effect of different degrees of reported planner or reviser styles did not vary across task execution, but a main effect of planner and reviser scores was established nonetheless: the higher the offline score, the more (or less) frequent do these activities occur during task execution.

The variation in distributions found for planning activities fits the available theory about the planner style. Students who report a higher degree of planner-type behaviour apply more planning activities at the start of task execution, but less towards the end of the writing process. This is in line with the idea that planners do most of their planning before they write anything down. The variation in distributions found for reading the assignment also fits the available theory about the reviser style. Students who report a higher degree of reviser-type behaviour read the assignment more often towards the end of task execution. This makes sense, because revisers think about a content plan after text production, that is, during later stages of task execution. It follows that typical revisers will also mostly think about the match between the produced text and the assignment during these later stages.

Various explanations come to mind for the fact that different distributions due to differences in reported writing styles could not be established for text production, reading own text and revising (i.e. there were main effects, but no interaction effects), and no effects were found at all for evaluating own text. One explanation is that, except for text production, these are low-frequent activities. The second explanation has to do with the nature of the activities in this specific age group. This seems to pertain particularly to text production, which is a frequent activity. Although its probability of occurrence is different in different episodes, this variation between different moments in the writing process is quite a bit smaller than it is in more developed writers. (Cf. Van Weijen (2009), who used the same assignments in a group of first year university students and found that text production activities were not likely to occur at all at one stage of task execution, but very likely to occur during other stages. Temporal variation was, in short, much larger.) There is, in other words, less variation to explain in the first place.

Cromley and Azevedo (2006) and Veenman et al. (2003) found that offline reports had little or no predictive value for online task execution. In the present study, relations between self reports and online task execution have been established, as is the case in the study by Torrance et al. (1999). There are two main differences between these two sets of studies which may explain the different findings. The first is that in the present study and the study by Torrance et al. (1999), the data are analyzed temporally, which is not the case in the studies by Cromley and Azevedo (2006) and Veenman et al. (2003). However, the absence of a temporal analysis in the former studies cannot be the sole explanation for the absence of relations between offline and online data. First of all, the fact that frequential main effects were found in the present study for text production, reading the assignment and revising, demonstrates that a temporal analysis is not always needed. In addition, Cromley and Azevedo (2006) and Veenman et al. (2003) studied reading tasks, whereas the present study and the study by Torrance et al. (1999) deal with writing processes. The reported writing styles—planner and reviser styles—imply variation in distributions, whereas this is

not so much the case for the offline measures used by Cromley and Azevedo (2006) and Veenman et al. (2003). Whereas there is evidence available that the occurrence of cognitive activities can also vary across the reading process (Janssen et al. 2005), this is not what offline measures of reading tasks generally focus on. It follows that the predictive value of these particular offline measures cannot be analyzed temporally.

It is striking that there were fewer effects—namely three—due to reported reviser behaviour than due to reported planner behaviour—namely five. Although it is possible that this is a chance finding, taken together with the fact that the reviser dimension in the Writing Style Questionnaire had lower reliability than the planner dimension (.64 versus .72), this seems to raise some doubts as to the usability of the reviser dimension for less proficient writers, such as the participants in the present study. This idea is supported by the observation that revising, and also reading and evaluating own text, which are associated activities, are extremely low-frequent activities (in this age group). In addition, it might be the case that the definition of revisers is not that clear-cut. It seems that two definitions are simultaneously in operation: one which focuses on the tendency to rely on revision, and one which focuses on how revisers use text production as a means to arrive at a content plan. Actually, the tendency to revise might be a side-effect of revisers' use of text production to get an idea of what they want to say. After all, their initial text production serves planning purposes and the resulting text is therefore likely to need some work. Possibly, the items in the Writing Style Questionnaire which deal solely with the amount of revision need reconsideration, as these might not be central to the definition of a reviser writing style. An example of such an item would be this statement: "When my text is ready, I elaborately read through it and make improvements: a lot can still be changed at that point". A Writing Style Questionnaire item which typically represents the part of the definition focusing on using text production to construe a content plan is: "For me, writing is a way to get my thoughts clear". On the basis of the present results, then, it seems than the planner dimension of the Writing Style Questionnaire can better predict different online configurations than the reviser dimension.

In this study the planning activity was construed of five subcategories (monitoring, goal setting, generating, structuring, metacomments). In future research, however, the validity of the analysis could possibly be increased by modeling the occurrence of these subcomponents separately. Hayes and Nash (1996), for example, distinguish between 'content planning' and 'non-content planning'. Goal setting, generating and structuring might arguably be instances of content planning, whereas monitoring and metacomments are more process-oriented activities and might therefore be seen as instances of non-content planning. Ideally, the relation between reported planner style and online planning activities should be analyzed separately for different types of online planning. This was not possible in the present study, due to the low frequency with which the subcomponents occur.

To conclude, it seems that questionnaires can have predictive value for online task behaviour. Kieft et al.'s (2006, 2008) Writing Style Questionnaire, and particularly its planner scale, seem to be a valid predictor of writing processes. In addition, a temporal analysis of (meta)cognitive activities across task execution seems to be a valid and sensitive reflection of online processing, particularly for writing. Whether a temporal analysis is also suitable for bringing out relations between offline and online measurements for other types of tasks, such as reading and mathematic tasks, is an issue for future research.

Appendix A

Example of an assignment.

Surveillance cameras in inner city areas

NACP, the National Action Committee for Pupils, is organising a national essay contest, especially for pupils of your age. You're also taking part. You absolutely want to win. The winning essay will be printed in PAUZE, a monthly magazine that is read by pupils your age from all over the Netherlands.

The subject of the essay has already been decided and was described in PAUZE 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. NACP is going to pay attention to this subject in a special edition of PAUZE. We want to hear from pupils 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, fellow pupils, 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 reporterd 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

Let A_{ij} be the proportion of an activity of writer j (j=1, 2, …, 20) at episode i (i=1, 2, …, 5), and P_j and R_j the planner and reviser scores of this writer. The model used to analyse the data can be written as:

$$Logit(A_{ij}) = \beta_0 * epi_{ij}^0 + \beta_1 * epi_{ij}^1 + \beta_2 * epi_{ij}^2 + \beta_3 * P_j + \beta_4 * epi_{ij}^1 * P_j + \beta_5 * R_j + \beta_6 * epi_{ij}^1 * R_j$$
$$+ \left[u_{0j} + u_{1j} * epi_{ij} \right].$$

Three notes:

1. Note that epi^0=1.
2. Remember: Logit (Y) = Ln (Y/1−Y).2. The dependent variables are proportions. As both the means and their variances are estimated (i.e. the variance of the residuals), in effect the proportion for each individual is estimated. Therefore, the intra-individal variance is not estimated. Only the variance of the between-writer residuals (i.e. u_{0j} and u_{1j}) is estimated.
3. Episode^2 was not included as a term for all of the six activities. This was, as explained in the 'Method' section dependent on whether lower-order terms have reached significance and whether the term itself makes a significant contribution to the model.

Appendix C

Parameter estimates for the average distributions for all of six (meta)cognitive activities (β_0, β_1, and, β_2), as well as the variances (S^2_{u0j} and S_{u1}^2) and the respective standard errors (between brackets). Note that 'episode' was re-scaled around the mean episode (= episode 3).

If the ratio of a fixed parameter exceeds 1.96 times its associated standard error, it is assumed to contribute significantly to the fit of the model. For random parameters, a more relaxed criterion for significance is used. After all, variances cannot take negative values. Therefore, we are dealing with a one-sided hypothesis being tested. A variance estimate contributes significantly if it exceeds 1.65 times the associated standard error. So, in effect a 5% significance level is in operation for both fixed and random parameters.

All estimates in this table are significant.

Activity	Fixed parameters			Random parameters	
	β_0 *epi^0	β_1 *epi^1	β_2 *epi^2	S^2_{u0j}	S^2_{u1j}
Reading assignment	−2.91 (.09)	−.31 (.04)	.04 (.02)	.47 (.09)	.07 (.02)
Planning	−3.00 (.11)	.25 (.03)	.04 (.02)	.77 (.14)	.03 (.01)
Text production	−0.37 (.04)	.01 (.00)	−.06 (.01)	.06 (.01)	.01 (.00)
Reading own text	−3.69 (.12)	.28 (.04)	.06 (.02)	.78 (.15)	.09 (.02)
Evalating	−5.38 (.17)	.38 (.09)	.09 (.05)	.70 (.22)	.28 (.09)
Revising	−8.16 (.58)	.25 (.02)	.26 (.08)	4.22(1.67)	1.62 (.62)

These estimates are expressed in logits. How to read and interpret them is described below. The activity 'reading the assignment' is used as an example.

As the 'episode' variable was centred, episode 1 became episode −2, episode 2 became episode −1, episode 3 became episode 0, et cetera. For reading the assignment, then, the (mean)

logit in the first episode equals $\left[-2.91* - 2^0 + -0.31* - 2^1 + 0.4* - 2^2 =\right] - 2.13$. When converted to proportions this equals $[\mathrm{expo}(-2.13)/(1 + \mathrm{expo}(-2.13)) =]0.10$ (see also Fig. 3).

The variance between writers is also significant for reading the assignment. The intercept variance equals 0.47. Hence, in episode 3 (re-scaled as episode 0) the 80% (z score $=1.28$) confidence intervals for the observed proportions varies from $[\mathrm{logit}(-2.91 - \sqrt{0.47}*1.28) =]0.05$ to $[\mathrm{logit}(-2.91 + \sqrt{0.47}*1.28) =]0.22$. Writers also differ from each other with regard to the linear change of reading the assignment. This means that the 80% confidence interval for the linear change varies from $[\mathrm{logit}(-.31 - \sqrt{.07}*1.28) =] - 0.61$ to $[\mathrm{logit}(-.31 + \sqrt{.07}*1.28) =]0.26$. Whereas for most students there is a clear decrease in the proportions of reading the assignment, some students show an increase in this activity during writing.

The parameter estimates in logits are somewhat hard to interpret. Therefore the mean changes over the five episodes are presented in proportions in Fig. 3. Episode numbers were, for the sake of ease of interpretation, also reconverted into their original numbers.

Appendix D

Variation in distributions due to Writing Questionnaire Scores. Parameter estimates and standard errors (between brackets) in logits. P: Planners core; R: Reviser score

Activity	Fixed parameters			
	$\beta_3 * P_j$	$\beta_4 * P_j *\mathrm{epi}_j{}^1$	$\beta_5 *R_j$	$\beta_6 * R_j *\mathrm{epi}_j{}^1$
Reading the assignment	−.227 (.09)	−.012 (.038)	.105 (.086)	.074 (.037)
Planning	−.150 (.065)	−.100 (.030)	.061 (.031)	−.014 (.009)
Text production	.087 (0.031)	.002 (.014)	.040 (.031)	−.006 (.015)
Reading own text	−.082 (.104)	.060 (.047)	−.252 (.106)	.009 (.048)
Evaluating	.172 (.128)	.044 (.077)	.021 (.138)	.003 (.084)
Revising	1.181 (.305)	.189 (.232)	.535 (.472)	−.055 (.327)
	Random parameters			
Reading the assignment				
	β_0		β_1	
β_0	.428 (.081)		–	
β_1	.115 (.030)		.063 (.016)	
Planning activities				
	β_0		β_1	
β_0	.789 (.140)		–	
β_1	.043 (.027)		.028 (.009)	
Text production				
	β_0		β_1	
β_0	.052 (.011)		–	
β_1	.001 (.004)		.008 (.002)	
Reading own text				
	β_0		β_1	
β_0	.601 (.119)		–	
β_1	−.020 (.038)		.092 (.024)	

Evaluating own text

	β_0	β_1
β_0	.613 (.202)	–
β_1	.066 (.087)	.199 (.074)

Revising

	β_0	β_1
β_0	1.320 (.968)	–
β_1	−.442 (.577)	1.301 (.623)

References

Biggs, J., Lai, P., Tang, C., & Lavelle, E. (1999). Teaching writing to ESL graduate students. A model and an illustration. *The British Journal of Educational Psychology, 69*, 293–306.

Breetvelt, I., Rijlaarsdam, G., & Van den Bergh, H. (1994). Relations between writing processes and text quality. When and how? *Cognition & Instruction, 12*(2), 103–124.

Cromley, J. G., & Azevedo, R. (2006). Self-report of reading comprehension strategies. What are we measuring? *Metacognition and Learning, 1*, 229–247.

Flavell, J. H. (1979). Metacognition and cognitive monitoring. A new area of cognitive—developmental inquiry. *The American Psychologist, 34*(10), 906–911.

Flower, L. S., & Hayes, J. R. (1980). 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: Lawrence Erlbaum Associates.

Galbraith, D. (1996). Self-monitoring, discovery through writing, and individual differences in drafting strategy. In G. Rijlaarsdam, H. van den Bergh, & M. Couzijn (Eds.), *Theories, models and methodology in writing research* (pp. 121–141). Amsterdam: University Press.

Galbraith, D., & Torrance, M. (2004). Revision in the context of different drafting strategies. In G. Rijlaarsdam, L. Allal, L. Chanquoy, & P. Largy (Eds.), *Revision: Cognitive and instructional processes. Studies in writing, Vol. 13*. Dordrecht: Kluwer.

Goldstein, H. (1979). *The design and analysis of longitudinal studies. Their role in measurement of change*. London: Academic.

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, and applications* (pp. 1–27). Mahwah: 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: Lawrence Erlbaum Associates.

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

Healy, M. J. R. (1989). Growth curves and growth standards. The state of the art. In J. M. Tanner (Ed.), *Auxiology '88: Perspectives in the science of growth and development* (pp. 1–29). London: Smith-Gordon.

Janssen, T., Braaksma, M., Rijlaarsdam, G., & Van den Bergh, H. (2005). *Flexibility in reading literary texts; differences between weak and strong adolescent readers.* Paper presented at the 11th European Conference for Learning and Instruction. Nicosia, Cyprus, August 22–27, 2005.

Kieft, M., Rijlaarsdam, G., & Van den Bergh, H. (2006). Writing as a learning tool. Testing the role of students' writing strategies. *European Journal of Psychology of Education, 21*(1), 17–34.

Kieft, M., Rijlaarsdam, G., & Van den Bergh, H. (2008). An aptitude-treatment interaction approach to writing-to-learn. *Learning and Instruction, 18*, 379–390.

Lavelle, E., Smith, J., & O'Ryan, L. (2002). The writing approaches of secondary students. *The British Journal of Educational Psychology, 72*, 399–418.

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

Olive, T., Kellogg, R., & Piolat, A. (2008). Verbal, visual, and spatial working memory demands during text composition. *Applied Psycholinguistics, 29*, 669–687.

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, and applications* (pp. 107–125). Mahwah: Lawrence Erlbaum Associates.

Roca de Larios, J., Manchón, R., 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*, 30–47.

Russo, J. E., Johnson, E. J., & Stephens, D. L. (1989). The validity of verbal protocols. *Memory & Cognition, 17*(6), 759–769.

Strömqvist, S., Holmqvist, K., Johansson, V., Karlsson, H., & Wengelin, Å. (2006). What key-logging can reveal about writing. In K. Sullivan & E. Lindgren (Eds.), *Studies in writing, vol. 18. Computer keystroke logging and writing: methods and applications* (pp. 45–72). Boston: Kluwer Academic Publishers.

Torrance, M., Thomas, G. V., & Robinson, E. J. (1994). The writing strategies of graduate research students in the social sciences. *Higher Education, 27*, 379–392.

Torrance, M., Thomas, G. V., & Robinson, E. J. (1999). Individual differences in the writing behaviour of undergraduate students. *The British Journal of Educational Psychology, 69*, 189–199.

Torrance, M., Thomas, G. V., & Robinson, E. J. (2000). Individual differences in undergraduate essay-writing strategies. A longitudinal study. *Higher Education, 39*, 181–200.

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, and applications* (pp. 207–232). Mahwah: Lawrence Erlbaum Associates.

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., Janssen, T., Braaksma, M., Van Weijen, W., & Tillema, M. (2009). Process execution of writing and reading. Considering text quality, learner and task characteristics. In M. C. Shelley II, L. D. Yore, & B. Hand (Eds.), *Quality research in literacy and science education* (pp. 399–425). Dordrecht: Springer.

Van der Stel, M., & Veenman, M. V. J. (2008). Relation between intellectual ability and metacognitive skillfulness as predictors of learning performance of young students performing tasks in different domains. *Learning and Individual Differences, 18*, 128–134.

Van Weijen, D. (2009). *Writing processes, text quality, and task effects: Empirical studies in first and second language writing*. Utrecht: LOT Dissertation Series.

Van Weijen, D., Van den Bergh, B., Rijlaarsdam, G., & Sanders, T. (2008). Differences in process and process-product relations in L2 writing. *ITL Applied Linguistics, 156*, 203–226.

Van Weijen, D., Van den Bergh, H., Rijlaarsdam, G., & Sanders, T. (2009). Composing episodes. Exploring the hierarchical structure of L1 and L2 writing. Manuscript submitted for publication.

Veenman, M. V. J., & Spaans, M. A. (2005). Relation between intellectual and metacognitive skills: age and task differences. In *Learning and Individual Differences 15, 159–176*.

Veenman, M. V. J., Prins, F. J., & Verheij, J. (2003). Learning styles. Self-reports versus thinking-aloud measures. *The British Journal of Educational Psychology, 73*, 357–372.

Zohar, A., & David, A. B. (2009). Paving a clear path in a thick forest. A conceptual analysis of a metacognitive component. *Metacognition and Learning, 4*, 177–195.