How students perform synthesis tasks: An empirical study into dynamic process configurations

Van Steendam, E.; Vandermeulen, N.; De Maeyer, S.; Lesterhuis, M.; Van den Bergh, H.; Rijlaarsdam, G.

DOI
10.1037/edu0000755

Publication date
2022

Document Version
Final published version

Published in
Journal of Educational Psychology

AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Educational Psychology

ISSN: 0022-0663

https://doi.org/10.1037/edu0000755

How Students Perform Synthesis Tasks: An Empirical Study Into Dynamic Process Configurations

Elke Van Steendam[1], Nina Vandermeulen[1, 2], Sven De Maeyer[3], Marije Lesterhuis[3], Huub Van den Bergh[4], and Gert Rijlaarsdam[5, 6]

[1] Department of Linguistics, KU Leuven
[2] Department of Linguistics, University of Antwerp
[3] Department of Training and Education Sciences, University of Antwerp
[4] Department of Languages, Literature and Communication, University of Utrecht
[5] Department of Social and Behavioral Sciences, University of Amsterdam
[6] Department of Teacher Education, Norwegian University of Science and Technology

In this study we examine process configurations in synthesis tasks. We study whether these configurations are students traits or vary within students per task. In a national survey with a representative sample of 658 Dutch upper-secondary school students, we collected writing tasks, registered students' writing behaviors (via keylogging) and their task perceptions and assessed the quality of their texts. Each participant completed two informative and two argumentative synthesis tasks. Writing process configurations were based on a preselected set of writing behaviors that proved to be related to text quality: time spent on sources and production activities, switching between sources and between sources and text production, and speed of production; with reference to the phase in the process (first, mid, final part). Latent profile analyses distilled four process configurations, some of which were more likely to occur with the informative genre. One process configuration, that is, "Fast text production," was related to qualitatively higher text quality scores than the others. Additionally, at the age of 16–18 a writing process configuration is not a student trait: in most instances, we observed two or more task configurations within students. Writers' task experiences such as topic knowledge and topic interest predicted the occurrence of certain process configurations which could indicate adaptivity. The finding that writing configurations of writers vary even between similar tasks has important implications for the generalizability of (synthesis) writing research on the basis of a single writing task and process per student.

Educational Impact and Implications Statement

Students have many options to approach a source-based writing task. The predominant belief has been that a task approach is a student trait. This study, however, shows that 16- to 18-year-olds vary their approaches to both similar source-based tasks and source-based tasks differing in rhetorical aim (to argue or to inform). This has implications for writing research using a single task and for writing instruction and feedback on students' writing processes.

Keywords: genre, process configurations, synthesis writing, task adaptivity, writing process profiles

Supplemental materials: https://doi.org/10.1037/edu0000755.supp

A source-based writing task which has attracted much research attention recently is the synthesis task (Graves et al., 2010; Mateos et al., 2014; Raedts et al., 2017; Van Ockenburg et al., 2019). Synthesis writing tasks are so-called hybrid reading-and-writing tasks (Spivey & King, 1989). They require writers to read, reread, and process sources and select, integrate, and synthesize the source information into a new text, which should reflect an integrated and coherent overview and understanding of the source material. They

Elke Van Steendam https://orcid.org/0000-0002-1910-9634
Nina Vandermeulen https://orcid.org/0000-0002-6731-7470
Sven De Maeyer https://orcid.org/0000-0003-2888-1631
Marije Lesterhuis https://orcid.org/0000-0002-3808-556X
Huub Van den Bergh https://orcid.org/0000-0002-1320-5334
Gert Rijlaarsdam https://orcid.org/0000-0002-2633-7336
Nina Vandermeulen is now at Department of Language Studies, University of Umeå.

This research was supported in parts by a National Grant from NWO, the Dutch Research Council, Grant 405-14-301.

Data and R-scripts for all the analyses are available at the Open Science Framework (OSF) via https://osf.io/hu6rp/?view_only=2dd1c46e178a469da6754dae9a2fff87.

Correspondence concerning this article should be addressed to Elke Van Steendam, Department of Linguistics, KU Leuven, Faculty of Arts, Warmoesberg 26, 1000 Brussel, Belgium. Email: Elke.vansteendam@kuleuven.be

are considered important learning tasks in education because of their epistemic value: reading, rereading, integration, organization, and elaboration of different sources calls for knowledge transformation (Mateos et al., 2014). What is more, being able to select and integrate information from different sources into a well-structured new text is a crucial skill, also in the workplace (cf. Raedts et al., 2017).

In Search of Larger Functional Units

A small number of research studies have examined the hybrid reading-and-writing synthesis process. The majority of these studies focus on source-related activities because of their crucial importance to synthesis writing (Escorcia et al., 2017; Leijten et al., 2017, 2019; Vandermeulen, Van den Broek, et al., 2020). They find, for instance, that source use may be beneficial to text quality at the beginning of the writing process, but less effective at the end (Breetvelt et al., 1996; Vandermeulen, Van den Broek, et al., 2020). Writing process theory and research, however, show that actions (such as reading sources) are part of *strings* of actions (Baaijen & Galbraith, 2019; Baaijen et al., 2012; Hayes & Flower, 1980; Van den Bergh et al., 2016). One activity may activate another one or set in motion a string of actions and vice versa to the extent that activities may follow each other in a cyclical and iterative way. These patterns may also vary over time (Van den Bergh et al., 2016). Some combinations have a higher probability of occurring at the beginning of the writing process for example than in the middle or at the end where other strings of activities may occur. It is, therefore, important to study the interplay of activities, "larger functional units," and "related patterns of building blocks" (Van den Bergh et al., 2016, p. 68) because activities (or a series of activities) relate to each other in a means-end relation" (Rijlaarsdam & Van den Bergh, 1996, p. 121).

State-of-the-art writing process theory and research show that variation in patterns of activities explains text quality, especially when writing process phases are taken into account (Van den Bergh et al., 2016). It is thus not the distribution of "a single cognitive activity that contributes to the quality of the resulting text, but a certain combination at a certain moment in the process" (Van den Bergh & Rijlaarsdam, 1996, p. 96). Accordingly, text quality in synthesis tasks may be determined by the interaction between work in sources and text production activities (Martínez et al., 2015, p. 276) instead of by source-based activities alone.

Furthermore, variation in these writing process configurations can be the result of compensation (Rijlaarsdam & Van den Bergh, 1996; Van der Hoeven, 1997). A writer who has problems reading sources during the initial stage of the writing process, for instance, can compensate this by matching sources and text during the revision stage at the end of the writing process.

From the theory thus emerge two tenets for writing process research. First, writing activities need to be studied in relation to each other rather than in isolation. Second, the distribution of writing activities should be considered across the writing process. That is, we must attend to the moment in the writing process at which they occur and reoccur.

In reading research and research on source navigation, the search for larger, integrative, and holistic images of processes has become common (Karlsson et al., 2018; List & Alexander, 2017; Strømsø et al., 2020). Studies in this field statistically group similar configurations of cognitive and metacognitive activities and distill profiles from this data through cluster analyses, principal component analyses or latent profile analysis. In writing research, however, such studies are much rarer. They have their precursor in writing process research identifying writing styles or individual approaches with either questionnaires (Galbraith, 1992, 1999; Kieft et al., 2007; Lavelle, 2007; Tillema, 2012) or think-aloud data (Hayes & Flower, 1980; combined with keystroke logging in Levy & Ransdell, 1996). However, these studies did not quantitatively extract or distill profiles. The few writing research studies that extracted profiles have done so for undergraduate essay writing (Kim, 2020; Torrance et al., 1994, 1999, 2000) or graduate and faculty-level report writing (Van Waes & Schellens, 2003). In synthesis writing only a single study has statistically distilled process profiles (Escorcia et al., 2017). In all writing research "profile" studies, profiles were built either (a) via concurrent observational (Escorcia et al., 2017) or protocol data (with triple task methodology cf. Torrance et al., 1999), (b) retrospective questionnaire data (Torrance et al., 2000), or (c) textual data (i.e., planning sheets or revision changes from first to second draft) combined with video data of Internet source searching behavior (Kim, 2020). Only a single study built profiles on the basis of keystroke logging data (Van Waes & Schellens, 2003).

Process Configurations or Student Profiles?

The question remains, however, whether profile studies report process configurations or student profiles. After all, the majority of these studies administered a single task in a single genre per writer. However, process research shows that (configurations of) writing processes may vary depending on the genre (i.e., rhetorical purpose; Beauvais et al., 2011) and topic when genre is kept constant (Tillema, 2012; Van den Bergh et al., 2016; Van Weijen, 2009). For synthesis tasks, such a genre effect was also observed in a study in which topic was kept constant (cf. infra Vandermeulen, Van den Broek, et al., 2020). These findings indicate that, without the administration of multiple texts and genres per writer, it cannot be inferred if profiles are person-centered (i.e., personal approaches or student "profiles") or task-bound (i.e., task approaches). Therefore, it is unclear whether the results of writing profile studies by Torrance et al. (1994, 1999) and by Van Waes & Schellens (2003) generalize to individuals, profiles, or to writing processes configurations.

Notable exceptions are the studies by Torrance et al. (2000) and Levy and Ransdell (1996), which administered multiple tasks per writer and were, therefore, able to distinguish between process configurations or student profiles. In their multimethod study, Levy and Ransdell (1996) logged and recorded the (think-aloud) writing processes of a small group of writers ($n = 10$) for a number of weeks. The writers displayed a remarkably consistent allocation of writing activities across the writing process to the extent that the researchers used the term "writing signatures." "Like cursive signatures," these configurations of writing activities proved to be "distinctly different between individuals" and exhibited "small and unsystematic differences within individuals across time" (p. 158). Levy and Ransdell note that the writing process patterns were "neither inherent nor implied by any known theory of writing" (p. 158).

In contrast to Levy & Ransdell (1996), the profiles in Torrance et al. (2000) do not seem to be individual, person-specific but rather task-dependent. The researchers had a cohort of 48 psychology undergraduates write on average six essays during their 3-year degree course. After completing an essay students had to fill out a questionnaire on their strategy use during writing. On the basis of a cluster analysis four strategy profiles were extracted. Subsequently, the researchers tested for "within-writer consistency" (p. 188) by examining the "single most-used" strategy (p. 190) for each individual student, and the percentage of essays for which they had reportedly used that strategy.

Findings revealed that 63% to 71% of the essays students wrote were written with their "single most-used strategy" (p. 190). The researchers thus acknowledged a fair degree of within-writer consistency. At the same time, however, 30% of the students' essays were (reportedly) written using a different strategy. Consequently, "the majority did not, regardless of tasks, strictly adhere to a single strategy" (p. 190), which illustrates a considerable intra-individual variety in writing process configurations. Thus, even though Torrance et al. (2000) note that every writer had an "overarching method of working" (p. 182), they also emphasize that an intra-individual uniform or single approach to writing or strategy stability is "illusory" (p. 183).

In conclusion, there are several requirements for investigating writing process configurations and assessing whether these configurations are true "profiles." Specifically, each writer must engage in multiple writing tasks and the features of these tasks need to be randomized. The latter requirement also supports generalization of any findings. The two profile studies meeting these requirements present seemingly contradictory findings. These can, however, also be explained by a different methodology for studying process configurations (think-aloud vs. reported writing behavior) and sample size. More research, also on a larger scale, is needed looking into the issue, also for synthesis writing.

Process Configurations in Synthesis Writing Research and Factors That Affect Configurations

Exploring Task Configurations of Synthesis Writing: Single-Task Studies

Three synthesis writing studies provide us with insights into integrative process configurations: two fairly small-scale qualitative studies (Mateos & Solé, 2009; Solé et al., 2013) and a quantitative study (Escorcia et al., 2017). All three studies are based on one task per writer and on a specific task.

Mateos and Solé (2009) and Solé et al. (2013) conducted a qualitative interpretative event-based analysis of secondary-school students' observational audio- and video-data of reading and writing processes and their relation to text quality. Three profiles emerged from these studies with students positioned on a continuum from a recursive/elaborative/mediated writing process configuration to a more predominant reproductive, linear approach. Overall, the more recursive or elaborative process a writer displayed, the better the resulting synthesis task.

The only quantitative profile study in source-based writing is Escorcia et al. (2017) on college-level writers handwriting a synthesis text. The researchers used a hierarchical cluster analysis to extract three profiles from observational occurrence data of (meta) cognitive writing activities by 27 students, including source-related activities such as (re)reading or marking source texts, (re) reading of the draft, writing and editing activities. One profile displayed a synthesis writing approach characterized by extensive reading and source-related activities such as note-taking prior to writing followed by an interspersion of writing and reading both sources and synthesis text. These writers were called Precise Transcribers as "their approach to writing seemed to consist in registering key ideas from the source texts . . . and precisely transcrib[ing] these to construct the written text" (p. 258). A second profile consisted of fifteen writers without a clearly defined strategy. They seemed to write spontaneously without much source consultation and note-taking prior to writing. They also did not (re)read sources or text, nor did they edit during the writing phase. They were labeled Spontaneous Writers because they scored low on planning, source-related activities, reading and revising. Finally, five writers were identified as Active Revisers: they focused on erasing and rewriting (editing), copied less from the sources than the Precise Transcribers and paused frequently. These pauses seemed to correspond to moments of reflection followed by changes or completions in the text. The variation in approaches did not covary with text quality, although the Precise Transcribers performed less well on spelling and grammar than the Spontaneous Writers.

Although the researchers use the person profiles "Transcribers," "Writers," and "Revisers" to label the configurations, the latter should be seen as process configurations. We cannot infer that the participants will show the same profile when writing again (even in the same genre) with another topic and/or other sources. Thus, it remains to be seen whether the process configurations could be generalized across tasks or persons.

Factors That Affect Process Configurations

Task and person features can affect process configurations. Task features can, for example, be mode of writing and reading (paper vs. screen: Van Waes & Schellens, 2003), genre (Beauvais et al., 2011), or the complexity of relations between source information. Processes could also be influenced by characteristics of the particular individuals, such as reading comprehension and writing proficiency, and by individual interest in and content knowledge of the topic (Kellogg, 1987).

Task Factors. Vandermeulen, Van den Broek, et al. (2020) found genre-specific configurations of source-related activities for argumentative and informative synthesis tasks: in the argumentative genre, students switched more often between source texts and synthesis text in the middle of the process. However, the text quality of the genre was especially and positively influenced when students switched between sources and their own synthesis process-initially, for example, for note-taking and/or drafting. Spending too little or too much time in the sources process-initially and process-finally had a negative impact on text quality. For the informative genre, alternating between sources process-initially correlated with a higher text quality. Alternating too much or too little between the sources process-finally negatively impacted text quality. This scarce or excessive source-switching at the end of the writing process, as well as spending too much time in the sources process-finally, could indicate problems with the integration of information during production. Even though the research into genre effects on writing processes is scarce, the study above suggests that differences in process configurations and their relation with text quality are at least partly genre-specific. To investigate that, however, confounding genre and task effects need to be avoided. Therefore, researchers should administer multiple tasks and multiple genres per writer.

Although Vandermeulen, Van den Broek, et al. (2020) do not report process configurations but instead effects of single activities (albeit in different episodes of the writing process), their study does show that when examining holistic process configurations in synthesis writing, one must choose a specific genre, or manipulate that factor. That approach has not been chosen, however, in the profile or writing process configuration studies of the type discussed here.

Person-Related Factors. An important research question is whether intrawriter variability in process configurations, if observed, reflects adaptivity or instability. Do writers vary the configuration of their writing process activities according to personal characteristics, such as interest in the topic and prior knowledge, as could be expected of more proficient writers? Or does process variability indicate a personal struggle with the task or genre requirements, and a random rather than planned connection between activities? Previous research has established that more proficient and stronger writers may fluently and swiftly adapt their writing approach to a specific audience and task type (i.e., task demands) to maintain a specific writing standard. This may depend on their experience with a specific task or prior knowledge of a topic which may not be the case for less experienced and novice writers (Bereiter et al., 1988; Bereiter & Scardamalia, 1987; Hayes & Nash, 1996; Piolat, 1999; Van den Bergh & Rijlaarsdam, 1996). However, few studies investigate (the same) writers' adaptivity in writing approach and *process* across different tasks and genres, especially not on a large-scale. Such studies require multiple tasks and genres per writer, ideally also across time (cf. "longitudinal design" as pointed out by Allen et al., 2016).

Allen et al. (2016) and Snow et al. (2015) studied writers' adaptivity in text (writing *product*) characteristics. The researchers call adaptivity "flexibility" by which they mean the flexible, varied use of linguistic features across multiple similar writing tasks to "craft more effective text" (Allen et al., 2016, p. 913). Both studies show that writing proficiency is positively associated with linguistic flexibility across multiple narrative essays (coined the "Linguistic Flexibility Hypothesis"). When studying the 16 timed narratives written by 45 upper-secondary school students over 3 weeks, they showed that the more proficient writers were more flexible in adjusting the narrative and cohesive properties of their texts to the writing task's specific context. This was less or not at all the case for the less proficient writers.

It thus remains to be seen whether individual writers also vary in writing behavior, in process constellations, from one task to the other (for similar tasks or for different genres). If this is the case, we could ask whether this intrawriter *process* variability reflects flexibility or rather instability. Instability would then be reflected in significant variability in text quality scores. Allen et al. (2016), for example, also hint that linguistic *product* "flexibility may not [always] be a positive writing characteristic" (p. 921), that is, across different situations and/or genres. A large degree of intra-writer *process* variability reflecting instability across a number of similar writing tasks might (also) indicate learning and experimentation. Verheyden (2010), for instance, showed that text quality scores of beginning writers in grade 3 fluctuated from one similar task to the next ($n = 6$ tasks) in the course of one school year. The authors hypothesized that the individual variability in text quality performance signaled learning (cf. Rijlaarsdam et al., 2014). Thus, it seems probable that secondary school writers largely unfamiliar with the synthesis task in the current study explore the effectiveness of different approaches to the writing task and in the process are not always able to uphold the same standard of text quality.

The studies by Allen et al. (2016) and Snow et al. (2015) also show that linguistic flexibility was linked to prior knowledge and to reading comprehension (Allen et al., 2016, p. 920). For synthesis writing, Escorcia et al. (2017) also touch on the role that individual task representation and task perception could play in writers' writing strategies (p. 263). The researchers point to topic knowledge and disciplinary knowledge to explain some of their findings on "writing behavior dynamics" (p. 263). They refer to the way in which (i.e., the moment as well the intensity with which) the writers planned and read the sources. Indeed, relevant to the type of reading-and-writing task subject to investigation is the widely acknowledged influence prior knowledge has on reading comprehension of the sources as supported by data from several studies (Boscolo & Mason, 2003; Solé et al., 2013). Solé et al. (2013), for example, looked at the correspondence between prior topic knowledge of source texts and task performance and suggest that "a better performance is achieved when an acceptable level of prior knowledge is coupled with a more elaborative [that is, more recursive] task performance pattern" (p. 81). Hence, these studies highlight a potential interconnectedness between, on the one hand, task perceptions for instance as a result of topic knowledge and knowledge of task characteristics, and writing (and reading) performance on the other hand.

Some other person-related variables such as topic interest, motivation, or writing anxiety may also have an impact on the process configuration. The results in Torrance et al. (2000) suggest a weak relation between strategy and writers' subjective experiences of the writing (i.e., as something to be enjoyed or disliked). That is why a secondary aim of the current study is to investigate the relation between writing configurations and task- and person-related variables.

The Present Study

Three aims guided the present study, with subsequent research questions:

Generalizability of Process Configurations

The first aim is to examine synthesis composing process configurations, and their generalizability across genres and writers.

From the literature reviewed emerges a need for more large-scale empirical studies on effective configurations of writing activities for synthesis writing. Ideally, such studies include multiple writing process variables (*integrative* cf. Baaijen et al., 2012) across different phases of the writing process (*dynamic*) and are conducted with multiple tasks and genres per person (*generalizable*). In this study, we address that need by examining configurations of synthesis writing activities across different phases of the writing process.

The availability of multiple genres, texts, and processes per writer enables us to investigate (a) which process configurations represent the observed behaviors most adequately (RQ1A), (b) whether these configurations are independent of genre (generalizability across genres, RQ1B), and (c) whether a process configuration indicates a trait, a student *profile* (generalizability across tasks, RQ1C).

Relations Between Process Configurations and Their Output, Text Quality

The second aim is to study the effectiveness of process configurations, that is, their relation to text quality and its generalizability across genres. Synthesis writing studies have shown univariate relations between the frequency of a single process category, such as reading sources, with text quality, dependent on the moment in the process. The question, however, is whether differences in—multivariate—synthesis writing process configurations will also relate to text quality. In the literature, a claim is made that writing process configurations can be compensatory (Rijlaarsdam & Van den Bergh, 1996; Van der Hoeven, 1997; Van Weijen, 2009). In such a case, we would not expect differences between process configurations to covary with text quality. Therefore, we will investigate (a) whether process configurations covary with text quality (RQ2A) and (b) whether these relations are independent of genre (generalizability across genres, RQ2B).

Writer-Specific (Uni-)Configuration or Multiconfiguration of Processes?

If process configurations do not indicate a person-specific trait or profile (cf. RQ1C), the third aim is to explore factors that could explain this multiconfiguration of processes across tasks. If process configurations vary within writers, this may point to adaptivity. If adaptivity is in play, one would expect this to be a characteristic of the stronger writers, that is, of the writers producing the better texts (RQ3A). Writers who show multiple configurations across tasks may have set equal standards of quality across tasks. One would thus expect the writers to vary less in text quality (RQ3B). Within-student task process configuration variability coupled with significant variability in text quality may indicate instability rather than adaptivity. One would also expect that a task process configuration is partly determined by the interaction between writer (topic knowledge, experienced difficulty) and task as a sign of adaptivity. If adaptivity plays a role, variation in perceived task features such as topic knowledge, topic interest or experienced difficulty, may explain multiple process configurations within writers (RQ3C).

To investigate the three research objectives, we collected writing process and performance data of 658 upper-secondary school students writing four different synthesis tasks, two per genre, in a national representative sample totaling 2,310 task performances. The national sample itself is described in Vandermeulen, De Maeyer, et al. (2020). Writing process data were collected by logging and time-coding keystrokes, mouse clicks or movements related to the production of the text (Leijten & Van Waes, 2013). Data on the relation between task and writer (topic knowledge, effort, topic interest) were collected via a questionnaire after each writing task.

Method

Participants

This study contains data from a national representative sample, conducted in upper-secondary education in The Netherlands (January through February in Grade 12 and April through June in grades 10 and 11). More specifically, the sample represents the three final grades of the pre-university education track, the track that provides access to a study at university. In collecting the national sample, all required ethical standards were met, that is, (ethical) approval and active consent was obtained from all participating schools and students. A total of 658 students from 43 Dutch schools participated. These schools were randomly and proportionally to their size, selected with a two-stage cluster sampling method. Within each school, eight of 10 students per grade (24 of 30 students in total) were randomly selected to write four texts which resulted in a sampling standard error of .04 to .08 for an intraclass correlation coefficient (i.e., proportion of variance between schools) of 5% to 20%, respectively. Fewer students from the final grade were selected (see Table 1) as these students were preparing for central national exams and the school board did not want to add to their already considerable workload by having them participate in a national survey. A percentage of 62.4 (n = 428) female students participated in the national baseline.

(Data-Collection) Procedure

Equipment

Participants worked on laptops provided by the research team on which in addition to the writing task and sources (pdf), Microsoft office and Inputlog (Leijten & Van Waes, 2013) were installed. To produce a specific synthesis task (in Word), students had to first open the pdf-sources one by one. Inputlog records keystrokes, mouse clicks or mouse movements related to the production of the text. It time-codes these log data, needed to analyze the temporal dynamics of the writing behavior (Leijten & Van Waes, 2013). It logs a single Word-document and additionally records time in sources, as well as switches between them and the document.

Procedure

Data collection took place at each school during regular class hours. Per school a group of 10 to 20 students attended a 1-day data collection session (in a separate classroom with specific equipment) during which they wrote four synthesis tasks, two before and two after lunch. Two supervisors, researchers on the project and/or trained research assistants, guided and monitored data collection. They closely adhered to a protocol they had been

Table 1
Number of Students Per Grade and Per School

Grade	n Students	n Schools
Grade 10	270	34
Grade 11	271	35
Grade 12	117	13

trained in to safeguard fidelity of administration of the (sequence of) four data-collection phases.

1. Participants were briefed about the general goal of the study and the procedure. They then signed consent forms.

2. Supervisors then walked participants (10 minutes) through a set of technical instructions consisting of (a) a definition of a synthesis task and (b) a specification of two types of synthesis tasks followed by instructions (c) on how to deal with the source texts, (d) on the intended audience, (e) text length, (f) time, and (g) style (i.e., Use your own words and Avoid copy paste). The instructions had to enable students to write with a task representation in mind as the text type was not instructed in Dutch secondary education and was thus new to them.

3. Supervisors briefly trained participants in how to use the keystroke logging software Inputlog and then instructed participants to start with the first synthesis writing task. They were instructed to use the specific sources for the synthesis task; no other (online or offline) sources could be used. To capture the full writing process, students had to write in the Inputlog Word document. Hence, note-taking on paper or in a document other than the Inputlog document, for example, during source reading, was not allowed because this would not be logged.

Students had to start the Inputlog recording session once they started with the task, that is, right after the opening of the different sources, to avoid the logging of these actions. They got 50 minutes for the task after which they had to stop the Inputlog recording. A timespan of 45–50 minutes is in line with classroom practices in Dutch secondary schools where classes take 50 minutes. Additionally, vetting and piloting of the tasks confirmed that tasks were doable in 50 minutes. Data collection confirmed this: average time-on-task was 34.69 minutes (SD = 8.88 minutes). Students who finished earlier did filler tasks until 50 minutes had passed. After each synthesis task students filled out the same closed-ended questions probing for their task experiences. The whole group then took a short five-minute break before the students started with the second synthesis writing task. After lunch they wrote the other two synthesis tasks.

Instruments and Measures

Students each wrote four synthesis tasks (cf. the Writing Product section) during which their writing processes were being logged (cf. Writing Process section) and filled in a questionnaire subsequent to each task (cf. Task Questionnaire section).

Writing Product

Writing Task. For the synthesis task, students had to include and connect information from all available sources in a coherent way. For the verbatim instructions see Appendix A. Key to synthesis writing is the information integration process. Students are supposed to compare and contrast information from the available sources as to integrate the multiple perspectives in their synthesis text. Sources were newspaper articles, research reports, and infographics containing numerical information (one per task). Across tasks the sources' total number of words was kept more or less equal, although the number of sources varied per topic. Students were instructed to use all sources.

For reasons of generalizability (Van den Bergh et al., 2012), systematic variations of synthesis tasks were constructed, resulting in four topics, for which eight variations were created (see Appendix B for the construction scheme):

1. Four topics were selected corresponding to the study profiles in the upper grades of pre-university education: Economy and Society (topic: the pay gap), Culture and Society (topic: the human–wildlife conflict in Africa), Nature and Technology (topic: self-driving cars), and Nature and Health (topic: food additives).

2. Two genres were represented for each topic to generalize across rhetorical aims: argumentation for which students had to argue a specific position with information from all sources and an information synthesis task for which students had to present a state of affairs on a specific topic drawing on the sources.

3. Per topic–genre combination, four sets of sources were composed by crossing two variables: (a) the relation between the sources varied (contradictory/complementary) to generalize across information organization load; (b) the amount of irrelevant information in the sources varied (high vs. low) to generalize across information selection load.

This meant that also for the argumentative synthesis task for which students had to argue a specific position (i.e., pro or con a specific topic), they had to process both sources in favor and against that position and they had to consider the multiple perspectives to the issue in the case of contradictory source information.

Each student had to write two argumentative and two informative syntheses. For practical reasons, the task-sets were randomly assigned to students, keeping topics constant across the writing sessions. Across schools, however, the topic order varied at random to minimize order effects.

The task design reflects the variety in synthesis tasks and will thus make it possible to generalize our findings. While previous studies on source-based writing usually focused on a specific type of synthesis writing (e.g., an argumentative synthesis text based on sources with conflicting information), the aim of this study was to draw conclusions for synthesis writing in general.

Rating. The texts were rated with a holistic rating scale, accompanied by five benchmark texts, an assessment method which has been shown to result in high validity and generalizability (as evidenced in prior research: Bouwer et al., 2015; Tillema, 2012).

Instrument: Text Scale With Benchmark Texts per Genre. Our text scale consists of a set of five well-chosen texts, ordered according to text quality, covering the expected range of quality, with a rating scale (identifiable aspects of quality), accompanied by annotations on each of the texts, reflecting the aspects of the rating scale. For an example, see Appendix C. The selection of the five benchmark texts for each genre was based on the assessment of a subsample of texts, 150 argumentative and 150 informative synthesis texts, with Comproved, an online tool for comparative judgment (Lesterhuis et al., 2016). For more information on the construction of the rating scale see Vandermeulen, De Maeyer, et al. (2020). The benchmarks represented a range of text quality scores, with in the middle a text with a score arbitrarily set at 100, and four text scores one and two *SD* above and below the average benchmark. An *SD* was estimated at 25 points.

The benchmarks contained an annotated explanation of four quality aspects included in the holistic evaluation: of (a) information relevance and correctness, (b) integration of the sources into a new text with its own structure and overarching theme, (c) coherence and cohesion, and (d) language use. These quality aspects have been shown to be decisive quality criteria for synthesis writing tasks (Boscolo et al., 2007; Mateos et al., 2008; Solé et al., 2013).

Procedure. Sets of texts to be rated were assigned to a panel of overlapping rater teams (Van den Bergh & Eiting, 1989), each consisting of three raters. This rater design allows for generalization across raters. Raters scored the students' texts relative to the benchmark texts by comparing the students' texts to the five benchmarks, and then positioning these on the scale. All possible scores could be assigned (so also scores below and above the benchmarks).

Raters. A total of 42 trained teachers of Dutch, Master's students and graduate students familiar with the task type subject to study, were recruited as raters and compensated financially. Training involved an individual rating session of five texts with the rating scale (1 hr) and a subsequent online discussion session with two researchers on the project in panels of two to three raters, during an hour to an hour and a half. In line with the design of overlapping rater teams, randomly split subsamples of texts-to-be-rated were distributed across rater teams. As a result, every text was rated by three raters.

Writing Product Measures. Product measures are holistic text quality scores with a mean of 90.48 ($SD = 18.44$) for the argumentative genre and 90.46 ($SD = 18.44$) for the informative genre. Average jury reliability was .65, an acceptable level of reliability in student text rating (Jonsson & Svingby, 2007), especially because it is a generalizability coefficient across 48 juries (each time composed of three raters from the pool of 42) with minimal training. Variation between the 48 different juries is also small ($SD = .08$). The final score consisted of the average of the three separate scores of the three individual raters.

Writing Process

Preparing and Filtering the Behavior Data. Keystroke logging data collected with Inputlog were prepared by using (a) the time filter and (b) source recoding function. First, the writing processes were filtered at the last character typed. Students were instructed to type the sign "=" when they were finished writing and revising their text. In this way, we could filter out the actions logged by Inputlog after finishing the text that we considered "clutter" (e.g., closing the source texts and actions to stop the recording of Inputlog). Second, sources were grouped with Inputlog's source recoding function into either (a) the source texts writers had at their disposal, (b) the main synthesis task writers were working on, and (c) off-task sources. Off-task sources were excluded from the source analyses. Writing behaviors with more than 10% in off-task sources ($n = 67$ cases of 2,310 cases) were discarded.

Writing Behavior Measures. Inputlog provides more than 3,000 behavioral variables. Based on prior research (Vandermeulen, De Maeyer, et al., 2020; Vandermeulen, Van den Broek, et al., 2020), we selected behavioral measures for two aspects of the synthesis process, that is, source use and text production. Four criteria guided the selection of the behavioral measures for these two aspects.

1. Measures had been shown to be decisive for the synthesis writing process and/or product in prior empirical research.

2. Measures had to be easily accessible, clear, unambivalent, and straight-forward indices of a specific synthesis process or activity devoid of multi-interpretability, that is, representing clearly observable behavior. The choice for this criterion is also based in the educational aim of the study: behavioral measures should lend themselves for the development of targeted interventions for writing instruction or for process feedback purposes (cf. Vandermeulen, De Maeyer, et al., 2020; Vandermeulen, Van Steendam, et al., 2022).

3. In line with research and recommendations by Breetvelt et al. (1994), interval-related information for the behavioral measures had to be available reflecting measures' temporal allocation across the writing process.

4. The selected measures also had to be relative measures, calculated in proportions or actions per minute, allowing us to compare writing processes across writers (as time-on-task was not necessarily the same for all writers) and to increase generalizability.

Following these criteria, we selected five behavioral measures. Three of them reflect source use and two production, in principle for three intervals of equal duration (see Table 2). These three intervals largely represent the beginning, middle and end of the writing process. This three-interval division facilitates interpretation and transfer to an educational setting, and is in line with the prevalent practice in writing process studies adopting a functional dynamic approach (Kellogg, 1987; Van den Bergh & Rijlaarsdam, 1999; Van den Bergh et al., 2016). To further validate the choice, we additionally ran two sets of validation analyses. First, we replicated analyses of Vandermeulen, De Maeyer, et al. (2020) and Vandermeulen, Van den Broek, et al. (2020) to indicate which behavioral measures explained text quality and could also reveal differences between genres. Prior analyses (cf. Vandermeulen, Van den Broek, et al., 2020) showed genre differences in the (effective) configuration of a specific set of (especially source) behavioral measures. Second, measures also had to be mutually exclusive for the most part (that is, with correlation coefficients below .85) as evidenced by a correlation analysis (cf. Table D1 in Appendix D). The data-analyses conducted to validate the selection of measures for subsequent profile analyses, both multilevel models and profile analyses, are included in Appendix E.

Table 2
Overview of Selected Writing Behavior Measures

Process aspect	Behavioral measure	Interval 1	Interval 2	Interval 3
Source use	Proportion time in the sources = Time in sources	✓		✓
	Number of transitions per minute between the sources = Switches between sources	✓	✓	
	Number of transitions per minute between the sources and the synthesis text = Switches between sources and text		✓	
Production	Proportion of active writing time = Time writing	✓		✓
	Number of strokes per minute = Speed	✓	✓	

Note. ✓ = the measure for that specific interval is selected.

On the basis of these analyses, we selected those behavioral variables that explained text quality and varied between genres. Consequently, the variable Proportion time in the sources in Interval 2 as opposed to Intervals 1 and 3 was not included. In total, we included nine behavioral measures in the profile analyses compare Table 2, which lists both the original Inputlog measure names and their equivalents throughout the text.

Task Questionnaire

After completing each of the four writing tasks, participants responded to a short questionnaire about their task experiences. Four 5-point Likert scale questions (cf. Questions with descriptive statistics in Appendix F) directly reflect the experiences of the writer with the task in terms of topic prior knowledge and topic interest, and the writer's process experiences in terms of experienced task difficulty and the effort they put into the task.

Analyses

A visual inspection of histograms showed that the transition variables, more specifically *Number of transitions per minute between the sources* and *Number of transitions per minute between sources and synthesis text* were not normally distributed. Therefore, these variables were log-transformed to approach a normal distribution. Analyses were conducted with the log-transformed variables.

Process Configurations and Generalizability

To answer RQ1A and to analyze whether different writing process configurations can be distinguished from validated Inputlog scores, we performed a latent profile analysis (LPA). More specifically, we used a model-based cluster analysis approach (finite mixture model approach) to detect the optimum number of substantively meaningful process profiles in the population of writing processes (Harring & Hodis, 2016; Hickendorff et al., 2018). We use the term "profile(s)" here in its technical and statistical meaning as referring to process profiles extracted by an LPA. We shall subsequently refer to the process profiles as (process) configurations. The Inputlog scores were all standardized for the LPA. To determine the optimum number of process configurations we used Saaty's (1990) analytic hierarchy process (AHP), as implemented by Akogul and Erisoglu (2017), in which different information criteria (AIC, AWE, BIC, CLC, and KIC) on model fit are mathematically combined to counter subjective decision-making on one (or a selection of) different information criteria. The AHP methodology is a widely supported mathematical tool to support multicriteria decision-making. Deciding on the number of profiles to extract is a such a multicriteria decision-making problem as multiple information criteria can point to different conclusions (Akogul & Erisoglu, 2017).

Next, we estimated a mixed effects model for each writing behavior in which the scores are predicted by process profile membership, taking into account the complex structure in the data (writing process scores are the result of the combination of students, nested in schools, performing writing tasks). Estimated marginal means and pairwise comparisons based on these mixed models were used to test whether and how process configurations differed significantly on each of the writing behaviors.

To answer the question whether these process configurations can be considered genre-specific (RQ1B) we used multinomial generalized linear mixed effects models, using genre as a predictor for the probabilities to apply a certain writing process configuration. In these models we took into account the possible impact of individual writers on the writing process configurations by adding a random effect for student. A null model (Model 0) is compared with a model with genre as a predictor to test the effect of genre (Model 1).

To investigate variability or stability in writing process configurations in more depth for RQ1C and to determine whether the process configurations can indeed be labeled writer-specific process profiles or so-called "signatures" we tested whether process configurations are a variable or stable writer characteristic. To that end we constructed a variable on the level of the writer, counting the number of different writing configurations they displayed (nConfigurationsR), which will subsequently be referred to as the index of *writer process variability*. Four different scores were possible varying from a writer having used one type of configuration for all tasks to a writer having used four different writing configurations, one for each writing task.

Effectiveness

To investigate the relation between process configurations and product text quality for RQ2A on Effectiveness, we ran a mixed effects model with text quality as a dependent variable and cluster profile as a predictor. Writing behaviors/processes were classified in a process profile or configuration based on the most likely class given the posterior membership probability coming from the LPA. The dependent variable, text quality, was standardized. As individual text quality scores are the result of the combination of students, nested in schools, performing writing tasks, a mixed effects model is applied in which text quality scores are considered Level 1, students and tasks are crossed at Level 2, and students are nested in schools at Level 3. Two models were estimated: a null model without predictors (Model 0) and a model with configuration membership as a predictor (Model 1). To answer RQ2B, if process configurations, and their relation to text quality, are genre-specific, a third model was added (Model 2) with configuration membership and genre and the interaction between both as predictors.

Writer-Specific (Uni-)Configuration or Task-Related Multiconfiguration?

To get a grip on the meaning of writer process variability, we first investigated the relation between writer process variability (cf. Process Configurations and Generalizability) and text quality.

1. To explore whether writer process variability predicts text quality (cf. RQ3A), we added the writer process variability index to the best fitting and most parsimonious multilevel model predicting text quality (Model 1 or 2) resulting in Model 3.

2. In a second analysis we studied, conversely, if writers who exploit multiple process configurations also vary more in the quality of the texts they write (RQ3B). Therefore, we created an index of *writer product variability* reflecting the degree of variability in text quality for each writer. To that end we calculated for each individual writer the standard deviation of their text quality scores. We ran a regression model (generalized linear model) with the *SD* in text quality for each individual writer as an index of writer product variability as a dependent variable and the writer process variability index as a predictor.

Subsequently, we explored if students' perceptions for the specific task with regard to topic knowledge, topic interest, task difficulty and effort invested predicted the probability that a specific process configuration would occur (RQ3C). We ran multinomial generalized linear mixed effects models with knowledge, interest, effort and difficulty as predictors for the categorical variable "writing process configuration." Building further on the models for RQ1 (cf. Process Configurations and Generalizability), three additional models are estimated: a model with main effects of all four task experience variables (Model 2b); a model with interaction terms for all four task experience variables and genre (Model 3b); and a more parsimonious model with interaction terms (Model 4b).

All analyses were run in R (R Core Team, 2019), making use of the packages mclust (Scrucca et al., 2016), tidyLPA (Rosenberg, 2019), lme4 (Bates et al., 2015), mclogit (Elff, 2020), and emmeans (Lenth, 2020). Data and R-scripts for all the analyses are available at the Open Science Framework (OSF) via (https://osf.io/hu6rp/?view_only=2dd1c46e178a469da6754dae9a2fff87).

Results

Process Configurations and Generalizability

Process Configurations (RQ1A)

Rather than subjectively deciding on the number of profiles to extract, we use the AHP approach as suggested by Akogul and Erisoglu (2017). Integrating these fit indices according to the AHP approach into the *composite relative importance vector* (C-RIV; cf. Table D2 in Appendix D and code on OSF) shows that a model that estimates four distinct process configurations in the data has the highest C-RIV. Therefore, we can conclude that we can distinguish four distinct writing process configurations based on the Inputlog measures.

The four process configurations display differences in synthesis process "behavior." Table D3 of Appendix D lists per process configuration the estimated marginal means for each writing behavior and the significant pairwise comparisons to determine the extent to which the configurations differ from each other. Figure 1 displays the four different process configurations, each represented by a different color/symbol. It displays the pattern of distribution of the behavioral measures for source use and production across three intervals of the writing process for the four process configurations. It needs to be stressed that configurations are distilled from behaviors/processes per writing task and do not refer to writer profiles. At this stage we do not know yet whether the configurations are person-centered. The behavioral measures, represented by their respective centroids and confidence intervals, are expressed in *z*-scores meaning that a value of 0 for a specific measure should be interpreted as an average value (occurrence or number) for that specific measure. Values above and below 0 represent occurrences or numbers of a behavioral measure above or below average respectively.

When interpreting Figure 1 it needs to be emphasized that centroids with confidence bars always need to be interpreted vertically, that is compared with other configurations, instead of horizontally, from one interval to the next. Hence, when Process configuration 1 for example scores below average for the time spent in sources in Interval 3, this means that compared with all other configurations Process configuration 1 scores lower (or relatively low). It should not be interpreted as a significant decrease in the time spent in the sources from Interval 1 to Interval 3.

From Figure 1 we can infer that process configurations show significant differences in synthesis process behavior. First, we explain these differences per interval.

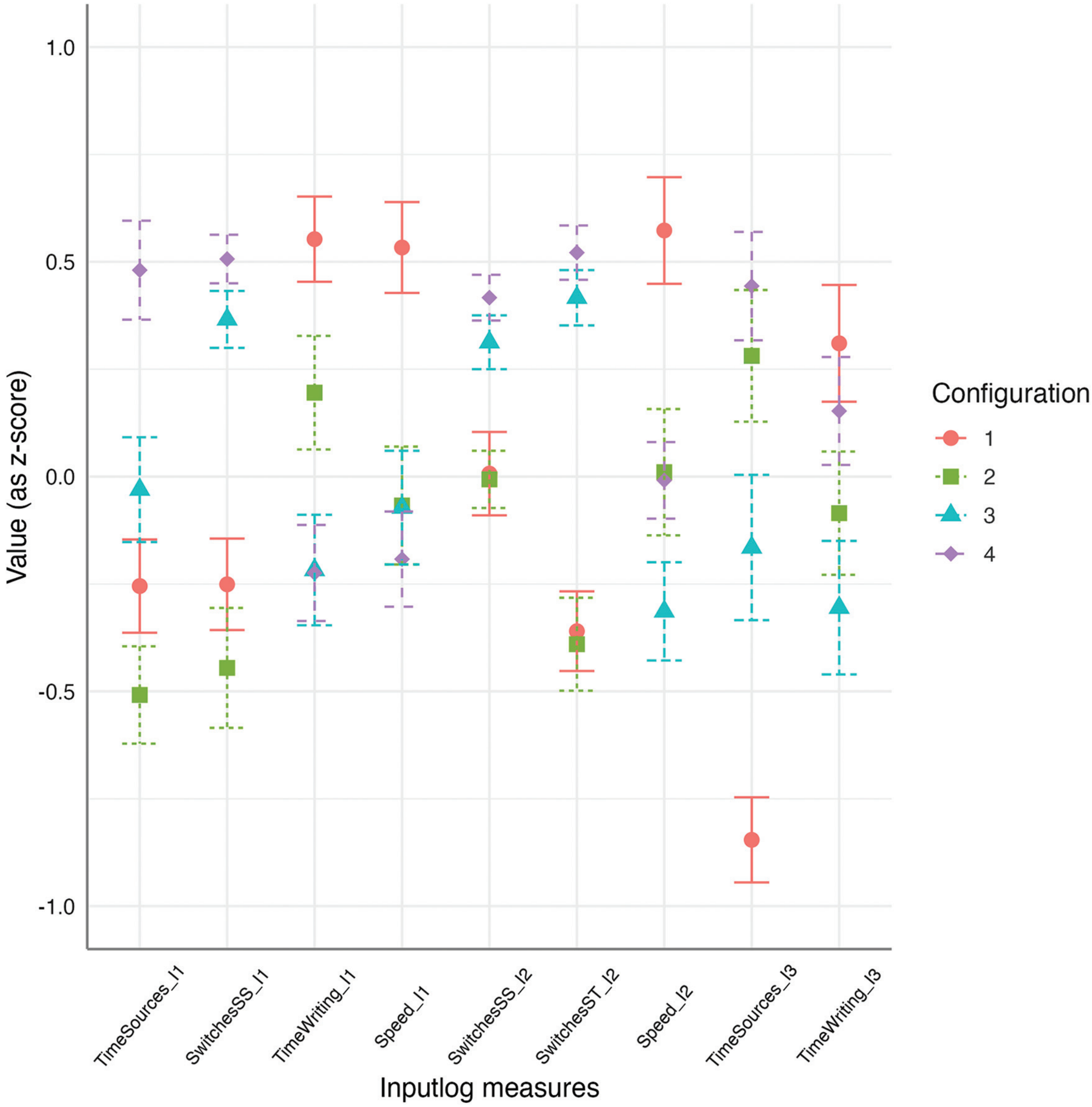
In the first interval, the beginning of the process, we see that Configuration 4 (in lilac/rhombus in all figures) stands out with an above average time in sources (to the left of Figure 1). This contrasts with the below average time in sources of Configuration 2 (in green/square). Both Configurations 3 and 4 (in turquoise/lilac or triangle/rhombus respectively) display an above average number of switches between the sources, while both configurations 1 and 2 display a below average number of source switches. Moreover, compared with the other process configurations, Configuration 1 (in orange/circle or dot) is characterized by a high writing time and a high writing speed.

In the second interval, Configurations 3 and 4 show an above average number of source switches, whereas Configurations 1 and 2 show an average number of source switches. Configurations 3 and 4 are also characterized by an above average number of switches between the synthesis text and the sources. For Configurations 1 and 2 this is below average. Regarding the writing speed in the middle of the writing process, Configuration 1 stands out with a relatively high speed.

In interval three, the last part of the writing process, Configuration 1 distinguishes itself from the other profiles with a below average time in sources. In addition, Configuration 1 shows the

Figure 1

Process Configuration Plot Showing the Average Scores for Behavioral Measures in z-Scores and 95% Confidence Interval per Process Configuration



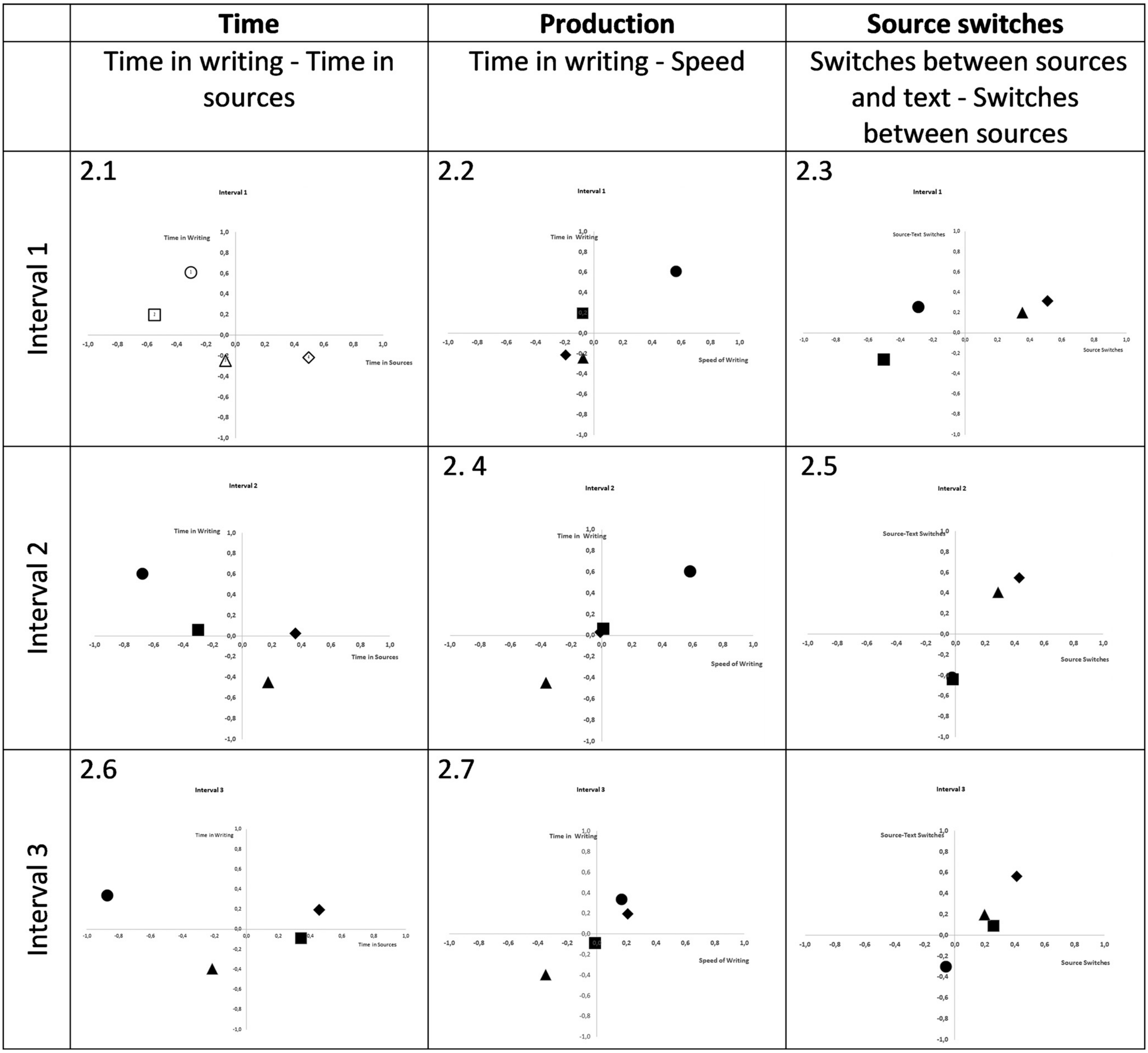
Note. Process configurations are represented by colored dots. Bars around each process configuration centroid represent respective 95% confidence intervals. TimeSources = Time in sources; SwitchesSS = Switches between sources; SwitchesST = Switches between sources and text; I1 = Interval 1; I2 = Interval 2; I3 = Interval 3. See the online article for the color version of this figure.

highest time in writing, while Configuration 3 shows the lowest time in writing.

For a more in-depth discussion and comparison of the process configurations across intervals, Figure 2 displays the process configurations along three dimensions. Five behavioral measures have been combined to present the configurations across (a) time with "time writing" (on y-axis) and "time in sources" (x-axis; left-hand panel), (b) Production, with "time" and "speed of writing" (middle panel), and (c) Source (use) behavior, with "switches between sources" and "between sources and synthesis text" (final panel). The measures for all three intervals (n = 15) are included of which only nine, decisive for the configuration comparison, were selected for the configuration analyses as reflected in Figure 1 (cf. Writing Behavior Measures).

Configuration 1 "Fast Text Production." Configuration 1 is characterized by a focus on text production (in terms of writing time and speed of writing) throughout the complete writing process (cf. Figure 1). Compared with the other configurations, the time in writing is high in the beginning and at the end of the writing process. At the end of the writing process, the time in sources

Figure 2
Time, Production, and Sources Switches in the Three Process Intervals for the Four Configurations



Note. Configuration 1 Fast text production = circle, Configuration 2 Last-minute source inspection = square, Configuration 3 Quick source use = triangle, Configuration 4 Source use = rhombus.

is low. There is also a high speed of writing in Intervals 1 and 2 compared with the other process configurations.

When describing the writing process characteristics for Configuration 1 (circle) over the three intervals, we see that students in this configuration spend a below average time in sources during the first interval (see the *x*-axis in Figure 2.1). They also switch between the sources below average (Figure 2.3). Their focus is on producing text as the speed of writing is well above average in the beginning of the process (Figure 2.2). This focus on text production is also present in the middle of the writing process as Configuration 1 is characterized by an above writing speed in Interval 2 (Figure 2.4). Both in the first and in the second interval, this configuration displays the highest writing speed of all profiles. Switches between synthesis text and sources are below average in Interval 2 (Figure 2.5). Given that at the end of the writing process (Interval 3), Configuration 1 shows a below average time in sources (the lowest of all configurations), and an above average time in writing, we can say that also in this part of the process the focus is on producing text (Figure 2.6).

Configuration 2 "Last-Minute Source Inspecting." Configuration 2 is characterized by an average text production (in terms of time in writing and speed of writing) in combination with a below average source behavior (in terms of time in sources and switches), except for the last interval as Configuration 2 shows an above average time in sources at the end of the writing process (cf. Figure 1).

When we look at the writing process interval per interval, we see that Configuration 2 displays both a below-average time in sources (square in Figure 2.1) and a below-average number of source switches (Figure 2.3) in the first interval. Compared with the other profiles, Configuration 2 shows the least attention to sources (cf. Time in sources) in the beginning of the process. In the same interval, writing time is slightly above average (Figure 2.2). During the second interval, the switches between the synthesis text and the sources are below average (Figure 2.5). In the last interval, Configuration 2 shows an above-average time in sources (though not as high as Configuration 4; Figure 2.6).

Configuration 3 "Quick Source Use." Configuration 3 is characterized by a high number of switches between the sources (in the beginning and middle of the process) and between the synthesis text and the sources (in the middle of the process) as illustrated in Figure 1. This above average switching behavior is combined with a slow text production (i.e., low speed of writing) in the middle of the writing process and little time in writing at the end of the writing process.

When describing the process in more detail, students in Configuration 3 (in Figure 2 in triangle) switch above average between the sources (though not as much as Configuration 4 in rhombus; Figure 2.3). Both the time in writing and speed of writing are below average (Figure 2.2). The second interval is characterized by the lowest writing speed of all configurations (Figure 2.4). The source switches and the source-synthesis text switches are above average (though, not as high as for Configuration 4; Figure 2.5). In the third interval, time in sources is slightly below average, and also time in writing is below average (the lowest compared with the other configurations; Figure 2.6).

Configuration 4 "Source Use." Configuration 4 stands out for its focus on source use (cf. Figure 1). Of all configurations, Configuration 4 shows the highest time in sources both in Interval 1 and 3, and the highest number of switches between sources in Intervals 1 and 2, and between synthesis text and sources in Interval 2.

When describing the synthesis behavior characteristics for Configuration 4 over the three intervals, we see that students in this configuration start their writing process by spending an above average time in sources (the highest of all configurations; rhombus in Figure 2.1). Both time in writing and speed of writing are below average at the beginning of the writing process (Figure 2.2). The switches between the sources are above average (and the highest compared with the other configurations; Figure 2.3). In the middle of the process, we see a combination of an above average number of source switches and synthesis text—source switches (both the highest of all configurations; Figure 2.5). Even in the last interval, Configuration 4 shows the highest time in sources of all configurations (Figure 2.6). The time in writing is slightly above average (Figure 2.6).

Generalizability

Latent profile analyses are run on a sample of texts and corresponding synthesis behaviors and distil different writing configurations from these writing behaviors. The issue now is whether these configurations are on the one hand genre-specific (RQ1B) and on the other hand writer-specific or person-centered profiles (RQ1C).

To answer the question whether configurations are genre-specific (RQ1B), a multinomial generalized mixed effect model is fitted with genre as predictor for writing process configurations. Comparing an unconditional model with a model including the effect of genre indicates that the latter fits the data best (loglikelihood ratio test: $\chi^2(3) = 74.16$, $p < .001$) leading to the conclusion that genre is related to the probability of using certain process configurations (cf. Appendix D, Table D4 for model comparisons). Figure 3 shows the predicted probabilities for each of the configurations as a function of genre (cf. Appendix D, Table D5 for the parameter estimates of this model). For informative texts there is a larger probability that writers engage in a process characterized by more source use (cf. Source use and Quick source use configurations). When writers write an argumentative text the different process configurations are somewhat equally probable.

The question whether the process configurations are participant-specific profiles (RQ1C) can be answered as well because participants wrote multiple texts. Profile analyses show that process configurations occur numerous times with the highest frequencies for both source-oriented configurations. Frequencies range from 771 times (Process Configuration 4 *Source use*) via 651 times (Configuration 3 *Quick source use*) to 409 and 405 (Configuration 2 *Last-minute source inspection* and Configuration 1, *Fast text production* respectively). The evidence suggests that many writers demonstrate more than one process configuration. The total frequencies of observed configurations exceeded the number of participants and writing tasks, which is only mathematically possible if some writers demonstrated multiple configurations across their different tasks. We created an index of writer process variability counting the number of times a writer displays a different configuration or process configuration. These analyses (cf. Figure 4) show that about half of the writers ($n = 305$, 49.4%) performed two process configurations across the four different synthesis tasks they executed, 150 writers (24.3%) three, and 151 writers (24.5%) showed only one process configuration. An almost negligible number of 11 writers (1.8%) displayed four process configurations.

We conclude that the observed process configurations are not identical with student traits. Instead, variability of process configurations within writers seems to be a (more) prominent characteristic with the majority of students showing (two to three) different process configurations and about 20% of the students showing four times the same configuration.

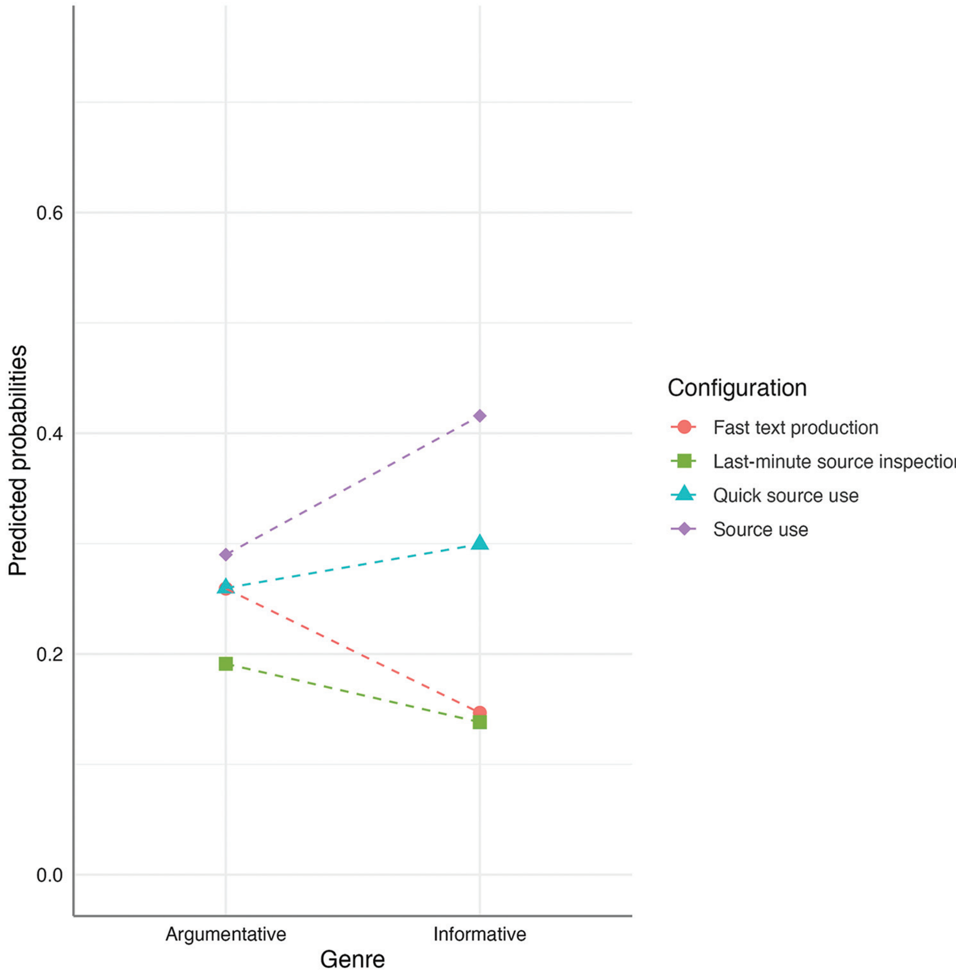
Effectiveness: Process Configurations and Text Quality (RQ2)

A multilevel model with the different process configurations (Model 1) shows a better fit than a model without the configurations to explain text quality (Model 0) (loglikelihood ratio test: $\chi^2(3) = 11.30$, $p = .010$). It also shows that especially Configuration 1, *Fast text production*, significantly predicts text quality. See Appendix D, Table D6 for parameter estimates of the fixed part of the model.

Configurations 2, *Last-minute source inspection*, and 4, *Source use*, score significantly lower on text quality than Configuration 1, *Fast text production*, as indicated by Tukey post hoc pairwise comparisons, $t(2022) = 3.227$, $p = .036$, Cohen's $d = .315$ and $t(2111) = 3.195$, $p = .020$, Cohen's $d = .312$, respectively). The Cohen's d values[1] for both differences point to a small effect. Previous research (Vandermeulen, De Maeyer, et al., 2020) showed a

[1] Cohen's d is calculated by dividing differences in text quality by the standard deviation for participants in the model, indicating the impact of Configurations on differences between participants.

Figure 3
Predicted Probabilities for the Four Configurations per Genre



Note. See the online article for the color version of this figure.

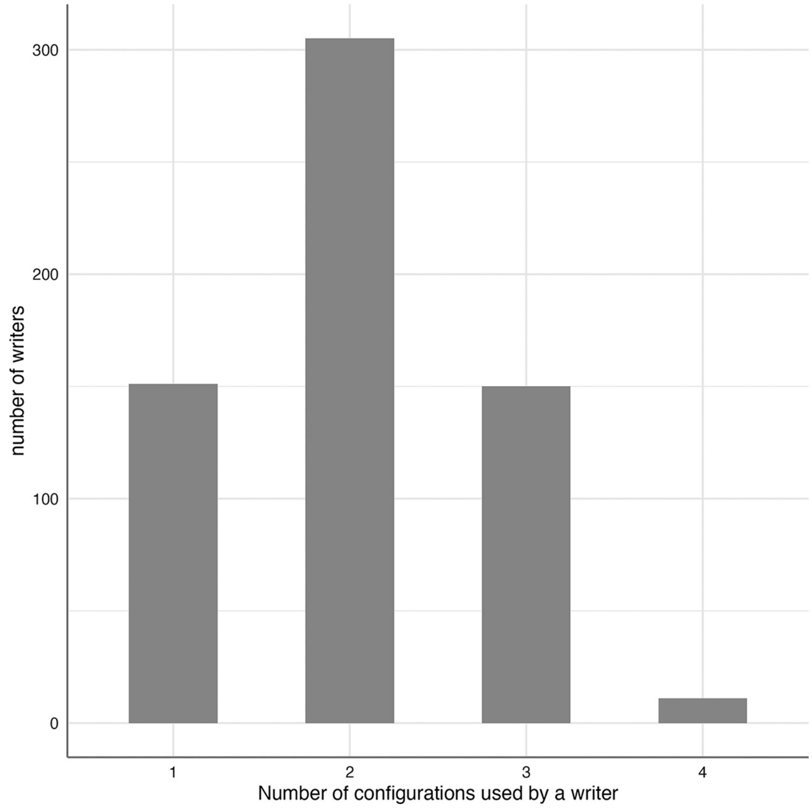
grade effect of 6 points on the same text quality scale with students of grade 12 outperforming students in grade 11 in their turn outperforming students in grade 10 for text quality. Using this as a point of reference, writing a text according to Configuration 1 results in a difference of approximately half a grade gain in text quality compared with writing a text according to Configuration 2 or Configuration 4. Other pairwise comparisons are statistically not significant. Parameter estimates for the different comparisons can be found in Table 3. This finding is generalizable across genres (RQ2B). Adding an interaction Configuration × Genre (Model 2) did not significantly improve the model (compare $\chi^2(4) = 1.131$, $p = .889$).

Adaptivity?

Adaptivity A (RQ3A)

Variation of process configurations within writers might point to task adaptivity, that is, to a writer adapting a process configuration to the task at hand, to produce the best possible text in different circumstances. If so, then variation in task configurations should have a positive effect on text quality. To analyze if this is the case, we added the index of writer process variability to the multilevel model with four different configurations (Model 1 cf. Section Effectiveness) resulting in Model 3 (cf. Section Writer-Specific (Uni-) Configuration or Task-Related Multiconfiguration?). Results (parameter estimates in Table 4) show no statistically significant effect of process variability, ranging from 2 to 4, on text quality. We can see that writers who each time deal with a synthesis writing task differently, that is, display four different configurations, score an average 3.86 points lower than writers who adhered to one and the same configuration for the four different synthesis tasks (Cohen's $d = .38$). However, the difference does not reach statistical significance, highly likely due to the limited number of writers who showed 4 different configurations ($n = 11$) and therefore the large standard error of the estimate. Running the model estimation without these 11 writers did not change any of the conclusions.

Figure 4
Bar Chart With Number of Different Process Configurations Used by Writers



We conclude that variation in the number of process configurations did not contribute to text quality scores: writers who vary more in their writing approach or process do not score better.

Adaptivity B (RQ3B)

A second perspective on process variability as a sign of adaptivity is that variation in behavior configurations may result in low variation in text quality within students as they try to optimize the writing process to keep up an internal standard of text quality. Therefore, we created a product variability coefficient (*SD*) ranging from 0 (no difference in text quality between the four texts written by the same writer(s)) to above 40, reflecting a high amount of text quality variability as shown in its distribution in Figure 5. Most of the writers as displayed on the *y*-axis in Figure 5 show a variability from 5 to 15 points on average. A generalized linear analysis with a gamma distribution and an identity link-function due to the right skewed distribution of the text quality variability coefficient shows that no statistically significant differences are observed in text quality variability (i.e., writer product variability) between the writers with regard to the number of process configurations they executed (i.e., from a single process configuration to four different ones as reflected in the writer process

Table 3
Post Hoc Comparison of Estimated Marginal Means for Text Quality per Configuration, Based on the Model Including Main Effects of Genre and Configuration on Text Quality (Model 1 cf. Analyses)

Comparison	Estimated difference	*SE*	*t* value	*p* value
Configuration 1 vs. Configuration 2	3.227	1.198	2.693	.036
Configuration 1 vs. Configuration 3	1.471	1.121	1.312	.555
Configuration 1 vs. Configuration 4	3.195	1.103	2.897	.020
Configuration 2 vs. Configuration 3	−1.756	1.089	−1.613	.371
Configuration 2 vs. Configuration 4	−0.031	1.061	−0.030	1.000
Configuration 3 vs. Configuration 4	1.725	0.927	1.861	.245
	Configuration 1	Configuration 2	Configuration 3	Configuration 4
Estimated marginal means	92.5	89.2	91.0	89.3

Note. Configuration 1 = fast text production; Configuration 2 = last-minute source inspection; Configuration 3 = quick source use and Configuration 4 = source use.

Table 4

Parameter Estimates (Est.), Standard Errors (SE), t Values, and p Value for the Estimates of the Fixed Part of the Model Including Main Effects of Process Configuration and Number of Distinct Process Configurations Used by a Writer on Text Quality (Model 3)

Measure	Est.	*SE*	*t* value	*p* value
Intercept[a]	92.767	1.796	51.653	<.001
Configuration 2	−2.869	1.205	−2.382	.017
Configuration 3	−1.499	1.124	−1.334	.182
Configuration 4	−3.250	1.109	−2.931	.003
Two configurations	−0.751	1.410	−0.532	.595
Three configurations	0.083	1.579	0.052	.958
Four configurations	−3.862	3.978	−0.971	.332

Note. Configuration 1 = fast text production; Configuration 2 = last-minute source inspection; Configuration 3 = quick source use and Configuration 4 = source use.
[a] Reference category: Configuration 1; one configuration.

variability index; cf. estimates in Appendix D, Table D7). The results do show, however, that writers who approach writing a synthesis task differently each time have a smaller intrawriter variability in text quality scores: their text quality scores vary less and are more stable. However, again probably due to the low number of writers in this category ($n = 11$), the result does not reach statistical significance ($t = .192$, $p = .847$). Running the model estimation without these 11 writers did not change any of the conclusions.

Adaptivity C (RQ3C)

In a final exploration to explain variability in writing process configurations, we relate task configurations to writers' task experiences, that is, (prior) topic knowledge, topic interest, experienced task difficulty and invested effort. We used multinomial generalized mixed effect models to analyze whether the probability to adopt a certain process configuration is a result of the score for each of the four perceived task experiences. Based on the model comparisons we learn that a model with two of the four task experiences, in interaction with genre, fits the data best (cf. Appendix D, Table D4 for model comparisons and Table D8 for the parameter estimates).

For topic knowledge we observe that there is a significantly different effect on the probability to engage in writing process Configuration 3, *Quick source use*, depending on the genre (p value = .0271). For argumentative texts topic knowledge is positively related to the probability to engage in *Quick source use* while for informative texts it is the opposite, the higher the topic knowledge the lower the probability to engage in *Quick source use* (see Figure 6). All other parameter estimates of the variable topic knowledge are not statistically significant.

Two parameters for the main effects of the variable topic interest are statistically significant (cf. Appendix D, Table D8). To aid the interpretation we plotted the predicted probabilities in Figure 6 (lower panel). Higher task interest increases the probability to engage in a *Source use* writing process or a *Fast production* process and decreases the probability to engage in a *Last minute source inspection* or *Quick source use* writing process. Given that none of the interaction terms are statistically significant we can conclude that these patterns are similar for both argumentative and informative texts.

Discussion

Discussion of Findings

This large-scale study has searched for synthesis process configurations, based on a national representative dataset of 658 secondary-school students responding to four tasks, two in an expository and two in an argumentative genre. Data were process behaviors, collected with keylogging software, text quality scores and students' perceptions of each of the four tasks in terms of topic knowledge, topic interest, effort and difficulty. With this data, this study addressed three main research objectives: (a) to examine synthesis writing process configurations, and their generalizability across genres and writers; (b) to study the relation between configurations and text quality and its generalizability across genres; and (c) to explore factors that could explain the possibility of configurations across tasks.

With regard to the first research objective (RQ1A), a latent profile analysis showed that four synthesis writing process configurations could be distinguished on the basis of an integrative and validated set of interval-based writing behaviors: (a) a Source use and (b) Quick source use process configuration characterized by a recursive and mediated source use, a (c) Fast production configuration which shows relatively speedy production and a (d) Process-final source inspecting configuration scoring low to average on both source use and production with a relatively heavy process-final source-orientation.

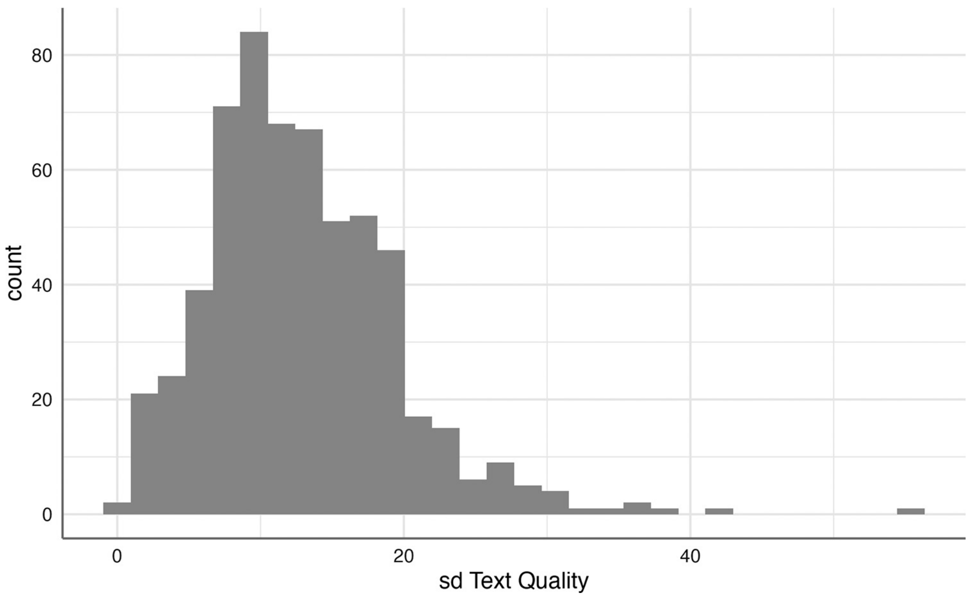
At least two of these process configurations reflect synthesis writing process configurations identified as student profiles by Escorcia et al. (2017) on the basis of observational data. Those two profiles are the Precise Transcribers and the Spontaneous Writers. Our Source users and Quick sources users, and Escorcia et al.'s Precise Transcribers share a relative, extensive pattern of source-related activities both prior to and during writing. Our Fast producers correspond to Escorcia et al.'s Spontaneous Writers because they show relatively few source-related activities. Compared with the Spontaneous Writers, however, they are avid and productive writers. They share some characteristics with profiles in profile research on other writing genres and task types; for example, with the Nonstop Writers of Van Waes and Schellens (2003) who score low for planning and revision and complete the writing quite quickly, or the Stage II Writers who do not pause a lot once they have started writing. However, they most resemble the Doers of Hartley and Branthwaite (1989) who studied the publication process and productivity of professional academic writers. To a certain extent, the Doers in their study scoring low on planning and revision are replicated in the Minimal Drafter by Torrance et al. (2000).

Whereas Torrance et al. (2000) could not statistically show an impact of a configuration on text quality, Hartley and Branthwaite's Doers did appear to be successful in their writing. Also, the Fast producers in the current study proved to be a highly productive and effective process configuration.

In fact, the texts resulting from the Fast production process configuration were the better texts (RQ2A). However, Hartley and

Figure 5

Intrawriter Variability in Text Quality Scores (Measured by the Standard Deviations) Across Numbers of Writers



Branthwaite's Doers were professional academic writers who may have adopted a specific routine for genres they were accustomed to writing. The secondary school writers in this study were relative novices to the synthesis task type which is currently not a standard component of Dutch secondary writing education (Van Ockenburg et al., 2019; Vandermeulen, De Maeyer, et al., 2020). Despite these very clear differences in writer population and tasks in both studies, there is a small point of comparison in that the Fast producers in this study may have been quite knowledgeable about the topic they had to write about. As a result, they may have felt quite confident in producing the task to the extent that a quick(er) inspection of the sources was sufficient. However, the analyses also show that the Fast production configuration does not have a higher probability of occurring as a result of a (higher) score on reported topic knowledge. Hence, instead of a possible effect of topic knowledge and/or consequent higher self-efficacy beliefs, it may simply be that the writers adopting the Fast production writing configuration, were (highly) proficient readers who as a result could afford to spend less time in the sources and focused on production instead. That the Fast text production approach is also a more likely route when writers are interested in the task (cf. results for RQ3C) may not only lend further credibility to the effectiveness of a Fast production configuration but also to the hypothesis that Fast text production is an approach adopted by the more engaged, interested, and proficient readers and writers. In any case, the Fast production approach paid off to the extent that writers who adhered to it wrote significantly better texts than both writers in the Source use configuration and the Last-minute source inspection configuration.

From a theoretical perspective on writing process research, three important issues emerge from these results. First, one type of behavior could signify two different functions. It seems to be fairly self-explanatory that to spend insufficient time in the sources process-initially and to refrain from comparing sources is not an effective strategy in terms of resulting text quality for this specific kind of synthesis task. Therefore, it is hardly surprising that the Process-final source configuration, which, compared with the other configurations, scores lowest for source use overall results in the poorest text (with the exception of the final phase of the writing process). However, for the Fast text production configuration, the same source use behavior leads to significantly better texts. Thus, what for some writers seems to be an indication of a poor organization of the writing process, for others appears to have the opposite effect, possibly as a result of proficiency differences for example, in reading (Boscolo & Mason, 2003; Kellogg, 1987; Solé et al., 2013). The results presented do not support that this may be the result of differences in topic knowledge.

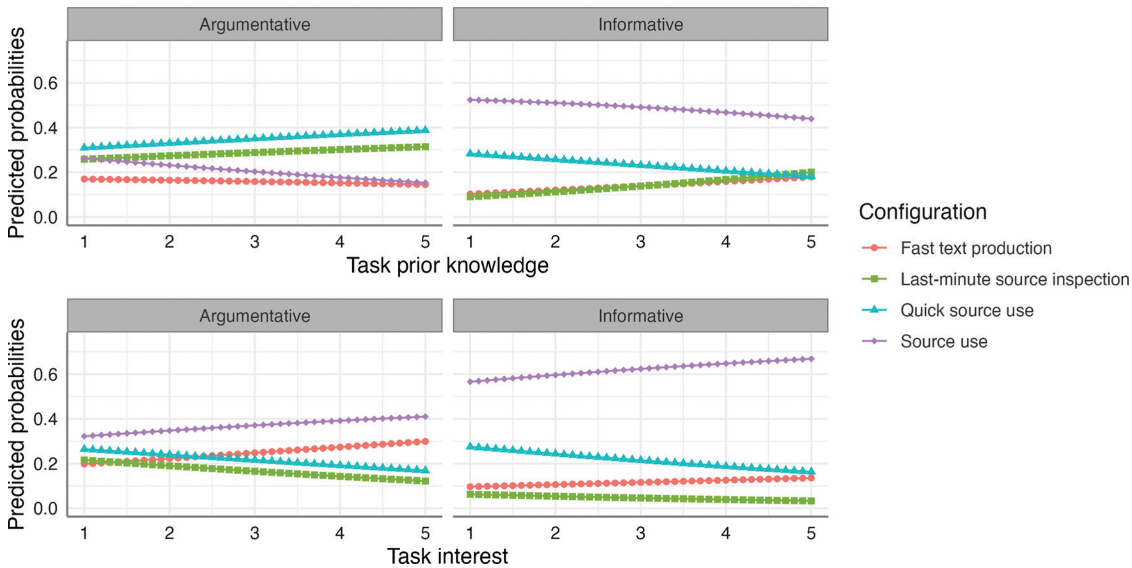
Second, the results also point out that types of activities do not stand on their own (cf. Van den Bergh et al., 2016). It is the way in which writing activities or process components interact that determines the effectiveness of writing processes. In the same vein as Van den Bergh et al. (2016), we argue that "it is not the single cognitive activity that contributes to the quality of the resulting text, but a certain combination, at a certain moment" (p. 68). Whereas Fast producers combine a low source use with avid production, Process-final source inspectors do not but instead turn to the sources at a moment in the writing process where that might be less effective.

Third, congruent with a functional dynamic view on writing process research, this last finding clearly illustrates that it is crucial when and to which extent an activity occurs in the writing process (Breetvelt et al., 1994).

Contrary to expectations, however, together with the Last-minute source inspection configuration, the Source use configuration seemed to be the least successful in terms of text quality. We follow Vandermeulen, Van den Broek, et al.'s (2020) line of reasoning in our interpretation of this finding. Given the importance of source-related activities for the synthesis task, it is hardly plausible

Figure 6

Predicted Probabilities Given Task Prior Knowledge (Upper Panel) and Task Interest (Lower Panel)



Note. See the online article for the color version of this figure.

to infer from this result that spending time in the sources and thoroughly comparing and contrasting them is not a sensible approach. On the contrary, in the same vein, we think this result could be explained by a nonlinear process-product relation. It is possible that too strong a focus on the sources has gone at the expense of production for the timed writing tasks in the current study. Alternatively, this strong focus may also be indicative of comprehension problems. However, latent profile analyses do not allow us "to capture more complex process-product relations by including curvilinear relations" (Vandermeulen, Van den Broek, et al., 2020, p. 256) as in LPA we only used the scores on the process measures as they were (and not in some kind of relation with text quality).

The results also provide us with empirical evidence that writing processes are genre-specific, which hitherto has only been investigated in a very limited body of research (cf. Beauvais et al., 2011; Vandermeulen, Van den Broek, et al., 2020) and mostly with a single task in a genre (= confounding task and genre effects; RQ1B). The results show that process configurations such as Source use and Quick source use with a heavy emphasis on source use are more likely to occur with informative synthesis tasks. This result is in line with genre differences in source use in Vandermeulen, Van den Broek, et al. (2020). The researchers showed that students displayed a similar source use behavior with regard to time spent in the sources and switches between them, across the writing process in an argumentative and informative genre. However, students writing an *argumentative* synthesis text switched more frequently between sources and synthesis text in the middle of the writing process (Interval 2). In the current study, a more intense switching between sources and synthesis text in Interval 2 combined with a lower time in the sources (across the entire synthesis writing process; i.e., Quick source use) is more common for the *informative* genre. Additionally, the Source use configuration, characterized by relatively much time in the sources, and switches between sources and between sources and synthesis text in the first and second interval of the writing process, is more probable for the informative genre.

Sample size and the inclusion of other process measures such as production in the current study could explain some of the differences with Vandermeulen, Van den Broek, et al. (2020). The findings, however, also raise several questions about the functionality of the process configurations. Do writers explore the sources more/longer or more intensely because the (expository) genre is less familiar and/or more difficult for them than the argumentative genre? Do these writers (feel they) lack an overarching organizing principle such as a standpoint as in the argumentative genre to guide source exploration? Or, alternatively, do informative synthesis tasks overall require longer, deeper, more thorough source exploration and comparison than argumentative synthesis tasks?

These speculative questions can only be investigated with think-aloud protocols, retrospective interviews or eye-tracking. The fact is that writers who know more about the topic they are writing an informative synthesis task about, are *not* more likely to display a Quick source use approach as is the case for writers with a high topic knowledge in the argumentative genre. Quick source use may thus serve a different function between the genres: whereas it seems to be a strategy for highly knowledgeable (topic) writers in the argumentative genre, the reverse is the case for the informative genre. Does this mean that for the latter, Quick source use signals (comprehension) problems with the task/sources or a coping strategy?

A possible genre-specific functionality, however, does not seem to be supported by a relation with text quality (cf. RQ2B). Contrary to Vandermeulen, Van den Broek, et al. (2020), the relation of the process configurations with text quality is not genre-specific, although the selected writing behavioral indices underlying the configurations (cf. Section Writing Process) were all selected based on a statistically significant relation to text quality which proved to be partially specific for genres (cf. Section Writing Process). Thus, that we could not replicate the genre effect to text quality for the distilled process configurations was an unexpected finding. It is possible that the effect of single variables is obliterated or obfuscated by compensating effects of other variables. Hence, whereas in the current study, we observed clear process configurations instead of person-specific profiles, the results also show that process configurations are genre-specific but their relation to text quality is not.

One of the significant findings to emerge from this study is that—in line with studies by Van den Bergh et al. (2012), and Van Weijen (2009)—we were "able to test for, rather than assume, within-writer consistency in writing behavior" (Torrance et al., 2000, p. 188). That we were able to do so with two genres and two tasks per genre is a strength of the study. The results show that most of the secondary-school writers did not adhere to a single writing configuration confirming that, depending on the specific task, writers organize their writing in a different way (RQ1C). The study thus adds to an increasingly growing body of research establishing intrawriter cross-task variability and, even more importantly, does so for synthesis writing configurations which have hitherto remained relatively unexplored, at least with different processes per student. This might be a step toward settling the issue of person-centered versus task-centered profiles, at the very least for synthesis writing. The research objective of this study was to investigate whether writers vary their writing configurations depending on task representation. We did not aim to closely scrutinize the impact of the different specific task characteristics such as number or type of sources (infographic vs. text) or the relation between these sources (complimentary/conflicting). Further research could study such task variability characteristics and their effect on writing configurations and more specifically, explore the interaction between task characteristics and writer characteristics for these writing configurations.

The main message to derive from these findings is that a fair degree of intra-individual cross-task variability exists in synthesis writing confirming the need for multiple tasks, *and* multiple processes, per writer to draw conclusions about writing proficiency or performance (Van den Bergh et al., 2012, 2016).

A question emerging from this intrawriter variability is whether it reflects adaptivity, leading to the optimal qualitative text, or rather instability, resulting in poor texts. We explored the question in two ways. First, we investigated the effect of the number of different writing process configurations per writer (intrawriter process variability) on text quality (RQ3A). Second, we explored the effect of this intrawriter process variability on intrawriter product variability (operationalized by the individual standard deviation in text quality; RQ3B). Results show no significant effect of variation in the number of process configurations on text quality. This means that for text quality scores it does not make a difference if writers deploy one, two, three or four different writing configurations. Neither could text quality variation be explained by the number of process configurations a writer adhered to. Both analyses confirm the same finding: in general, intrawriter variability has no impact on text quality. This result does not rule out, however, that for some writers there may have been adaptivity (e.g., to the task demands) whereas, for others, process variability may point to random writing process configuration choices (i.e., instability). A source of uncertainty for example remains for the writers displaying four different process configurations as they seem to score on average almost 4 points lower (about .20 ES) than relatively stable writers adhering to a single writing process configuration. Additionally, their text quality scores seem to be more stable (but remain lower in comparison). However, as both results do not reach statistical significance probably owing to the limited number of writers in the category that showed four process configurations, caution in interpretation and extrapolation must be applied.

We found indications that task perceptions affect the occurrence of a process configuration (RQ3C). As mentioned, whereas the effect for topic knowledge is genre-specific, the effect for topic interest is not. The higher the topic knowledge, the more a Quick source use configuration is displayed in the argumentative genre; the reverse is the case for the informative text genre. For both genres, a higher task interest increases the probability that writers engage in a Source use writing process or a Fast production process and decreases the probability that they engage in Quick source use or Last-minute source inspection. We know from the relation between process configurations and text quality for RQ2 that students who adopt a Fast text production configuration also write the better texts when compared with the Last-minute source inspection process configuration. Surprisingly, students who are more interested in the task are more likely to engage more with sources (Source use) but not necessarily effectively, to the extent that they write better texts as Fast text production also outperforms Source use. It might of course simply be an indication that topic interest even if combined with a more thorough source use is not sufficient to write a better text. These indications provide some depth to the interpretation of the process configurations. They are also a springboard for further analyses to explain intra-individual variation in process configurations and to study the interaction between writer and task variables that explain adaptivity.

The effect of topic knowledge on writing processes was demonstrated by Kellogg (1987). Based on two experiments with a between-subjects design, he observed that high topic knowledge reduced effort, but he could not demonstrate that topic knowledge resulted in differences in strategy allocation of planning, production and reviewing. His tasks were essay tasks, however, not source-based writing tasks. In the present study, with a within-subjects design, we found that topic knowledge affects the configuration of writing activities. The issue now is whether writers are aware of this and consciously adapt their strategy to the topic at hand. We demonstrated that writers vary their approaches and found that topic knowledge and topic interest can explain such variation, but we do now know whether this variation is strategic variation.

Opportunities for Future Research

This study may constitute a useful starting point for future research on the effect of person-related variables on (these) writing configurations. One of these variables could be motivation. Villalón et al. (2015) showed a link between self-efficacy beliefs, writing conceptions and synthesis writing quality. Cuevas et al. (2016) looked at the impact of transactional beliefs on collaborative synthesis writing. In future research in line with this study, we aim to investigate the relation between writers' motivation (self-efficacy, intrinsic and extrinsic motivation, learning goals), writing anxiety and transactional beliefs (White & Bruning, 2005) on the one hand and effective dynamic writing process configurations on the other.

Other individual characteristics worthy of examination are prior knowledge and reading comprehension (Plakans, 2009). In this study, topic knowledge was measured with a self-report measure. Prior research (as reviewed by Dochy et al., 1999) shows that in general, learners may be poor assessors of their own topic knowledge. The topic knowledge measure in the current study may have been more grounded as students were asked to assess how much they already knew about a topic after having written a synthesis task on it rather than prior to the writing task. Nevertheless, further studies including topic or prior knowledge *tests* could be carried out to study the relation between topic knowledge, topic interest and invested effort (cf. Kellogg, 1987) on the one hand and process constellations on the other. Another variable in need of further investigation is reading comprehension.

In fact, given the importance of reading for the hybrid reading-and-writing tasks, a potentially productive avenue for future research is to study process configurations from both a reading and writing perspective especially in the light of a state-of-the-art body of research on multiple document reading (Cerdán et al., 2018), processing of sources (Anmarkrud et al., 2014), and information selection (Cameron et al., 2017), also from a profile analysis perspective (cf. text navigation profiles by List & Alexander, 2017).

Additionally, future research could replicate this study for other populations such as university students or expand it to other, timed and untimed, source-based writing tasks such as academic literature reviews (Raedts et al., 2017). The results can also inform professional writing (e.g., reports) or multimodal writing for which writers need to process and integrate information from various audiovisual and written sources (Leijten et al., 2014; Raedts et al., 2017). Even though the generalizability of effective writing configurations for synthesis writing to other writing tasks or genres, also nonsource-based ones, was not the objective of this study, we believe the findings offer a theoretical and methodological surplus for writing profile research.

In this study, writers wrote four texts at roughly the same moment in time allowing us to investigate intrawriter variability. It of course also makes sense to study intrawriter variability as a result of maturation (in longitudinal profile studies such as Torrance et al., 2000) or learning, instruction and feedback as also pointed out by Allen et al. (2016) and Snow et al. (2015).

Given the sheer size of the national survey sample, we took our recourse to unobtrusive keystroke logging software to log students' writing processes. As a result, we were able to build process configurations on the basis of a "set of variables derived from the writing process itself" (Van Waes & Schellens, 2003, p. 849), reflecting directly observable synthesis writing behavior. One forte of the current study is that this set of variables is validated in multiple ways. First, variables selected are theoretically grounded and based on prior research for example, source use variables for a distinguishing feature of the synthesis task. Second, we took care to select relative interval-based writing behavioral measures congruent with a functional dynamic view on the writing process. Third, the variables selected all showed a statistically significant effect on text quality and a genre effect in separate multilevel models (cf. Appendix E). Additionally, the variables selected were mutually exclusive for the most part (cf. correlational analysis Table E1 in Appendix E).

However, use of keystroke logging also meant we had no immediately accessible information on planning behavior. Of course, pauses may offer information on cognitive processing and possibly on planning (Van Waes & Schellens, 2003). However, they are multi-interpretable and do not allow for an unequivocal interpretation of the cognitive processes involved, especially in the context of synthesis writing. Second, the selection criterion of interval-based process measures also meant that no revision measures were included as currently no automatically computed interval-based revision measures are available. Furthermore, the revision index calculated by Inputlog does currently not differentiate between copy paste behavior and pure revision behavior. Differences in revision behavior, however, clearly mark the profiles by Escorcia et al. (2017), Van Waes & Schellens (2003) and Torrance et al. (2000). Revision may have played a crucial role in the Process-final source inspecting configuration with its high amount of time in the sources at the end of the process and an average production throughout (compared with the other configurations). The low (standardized) scores for both source use and active writing time process-finally of the Quick source use configuration could perhaps point to revision. It can be expected that the addition of revision behavior to the profile analysis will allow us to interpret and further clarify the process configurations defined rather than result in completely different ones. We would also like to point out that the source use behavioral measures in Vandermeulen, Van den Broek, et al. (2020) already explained up to 24.6% of the variance in text quality.

On the basis of a relatively limited selection of nine theoretically grounded and empirically validated behavioral measures, the profile analysis in this study provides us with a varied, rich, and integrative window on configurations of writing behavior in synthesis writing and their effect on text quality.

Of course, a writing process is so much more than observable writing behavior; it involves goal setting, (strategic) decision-making and monitoring which cannot be adequately captured by keystroke logging measures. Research shows that aligning keystroke logging measures with (underlying) cognitive processes is complicated (Baaijen & Galbraith, 2019; Baaijen et al., 2012; Wengelin, 2006). That is why we carefully selected the behavioral keystroke logging measures and when interpreting the results, refrain from aligning writing behavior with (possibly underlying) cognitive processes.

This study has gone some way toward enhancing our understanding of advantages and limitations of keystroke logging measures for profile analysis. Future research could triangulate these measures with other methods or more fine-grained analyses for an unambiguous interpretation or alignment (cf. Baaijen et al., 2019) and for additional insight into strategic decision-making and regulation. Such methods could be screen recording (Knospe, 2017), eyetracking (De Smet et al., 2018), the trace reading methodology (Merchie & Van Keer, 2014), or even think-aloud protocols (Lindgren & Sullivan, 2019; Tillema, 2012). However, both their implementation and the required time-intensive analysis of the data collected make these methods less suitable for large-scale studies such as this one. For think-aloud protocols, for example, there is a potential, but contested, reactivity on writing behavior, especially in combination with keystroke logging (see also Baaijen & Galbraith, 2019).

Implications for Theory and Education

With regard to theory, as laid out in the literature review, profile studies in writing research are scarce. To the best of our knowledge, there is no all-encompassing theoretical writing model including cognitive, affective (interest, attitudes incl. motivation, dispositions and beliefs), and contextual variables (such as task and genre) underlying the existence of writing (task or person) profiles. Especially when adopting a functional dynamic approach to patterns of larger, functional writing process units "it is important to realize that the domain is still in the explorative, observational phase, providing us with correlational insights" (Van den Bergh et al., 2016, p. 69). There is, of course, important theoretical work (incl. writing models) in writing process research that can guide writing profile studies in general and the work subject to review in specific. In drawing on some of this theoretical work, we intend to contribute to theory-building, by empirically documenting an important methodological requirement for drawing meaningful conclusions on configurations or "profiles," that is, the administration of multiple texts and genres per students.

With regard to education, this study was part of a larger study to inform learners about their writing process, relative to the options possible. The findings of this study will be used to develop targeted interventions for writing instruction or for process feedback purposes. Instruction or feedback could exploit the different synthesis writing configurations to provide students with an insight into the writing process and into effective writing process configurations. Especially instruction or feedback aiming to expose students to different writing approaches by for example, having them contrast them or compare them to their own writing approach, could benefit awareness-building and reflection and ultimately result in learning.

Conclusion

In this study, we pursued a line of writing research that was started in 1980 when Hayes and Flower (1980) studied different process configurations on the basis of think-aloud data. In doing so, we believe to have contributed to a tradition of writing process typologies and research on writing configurations by showing that writing configurations are not person-centered or specific signatures but can vary within writers.

References

Akogul, S., & Erisoglu, M. (2017). An approach for determining the number of clusters in a model-based cluster analysis. *Entropy*, *19*(9), 452. https://doi.org/10.3390/e19090452

Allen, L. K., Snow, E. L., & McNamara, D. S. (2016). The narrative waltz: The role of flexibility in writing proficiency. *Journal of Educational Psychology*, *108*(7), 911–924. https://doi.org/10.1037/edu0000109

Anmarkrud, Ø., Bråten, I., & Strømsø, H. I. (2014). Multiple-documents literacy: Strategic processing, source awareness, and argumentation when reading multiple conflicting documents. *Learning and Individual Differences*, *30*, 64–76. https://doi.org/10.1016/j.lindif.2013.01.007

Baaijen, V. M., Galbraith, D., & de Glopper, K. (2012). Keystroke analysis: Reflections on procedures and measures. *Written Communication*, *29*(3), 246–277. https://doi.org/10.1177/0741088312451108

Baaijen, V., & Galbraith, D. (2019). Aligning keystrokes with cognitive processes in writing. In E. Lindgren & K. P. H. Sullivan (Eds.), *Observing writing: Insights keystroke logging and handwriting* (pp. 30–50). Brill.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Beauvais, C., Olive, T., & Passerault, J.-M. (2011). Why are some texts good and others not? Relationship between text quality and management of the writing processes. *Journal of Educational Psychology*, *103*(2), 415–428. https://doi.org/10.1037/a0022545

Bereiter, C., & Scardamalia, M. (1987). *The psychology of written composition*. Erlbaum.

Bereiter, C., Burtis, P. J., & Scardamalia, M. (1988). Cognitive operations in constructing main points in written composition. *Journal of Memory and Language*, *27*(3), 261–278. https://doi.org/10.1016/0749-596X(88)90054-X

Bergh, H., & Eiting, M. H. (1989). A method of estimating rater reliability. *Journal of Educational Measurement*, *26*(1), 29–40. https://doi.org/10.1111/j.1745-3984.1989.tb00316.x

Boscolo, P., Arfé, B., & Quarisa, M. (2007). Improving the quality of students' academic writing: An intervention study. *Studies in Higher Education*, *32*(4), 419–438. https://doi.org/10.1080/03075070701476092

Boscolo, P., & Mason, L. (2003). Topic knowledge, text coherence, and interest: How they interact in learning from instructional texts. *Journal of Experimental Education*, *71*(2), 126–148. https://doi.org/10.1080/00220970309602060

Bouwer, R., Béguin, A., Sanders, T., & van den Bergh, H. (2015). Effect of genre on the generalizability of writing scores. *Language Testing*, *32*(1), 83–100. https://doi.org/10.1177/0265532214542994

Breetvelt, I., Van den Bergh, H., & Rijlaarsdam, G. (1994). Relations between writing processes and text quality: When and how? *Cognition and Instruction*, *12*(2), 103–123. https://doi.org/10.1207/s1532690xci1202_2

Breetvelt, I., Van den Bergh, H., & Rijlaarsdam, G. (1996). Rereading and generating and their relation to text quality: An application of multilevel analysis on writing process data. In G. Rijlaarsdam, H. Van den Bergh, & M. Couzijn (Eds.), *Theories, models and methodology in writing research* (pp. 10–21). Amsterdam University Press.

Cameron, C., Van Meter, P., & Long, V. A. (2017). The effects of instruction on students' generation of self-questions when reading multiple documents. *Journal of Experimental Education*, *85*(2), 334–351. https://doi.org/10.1080/00220973.2016.1182884

Cerdán, R., Candel, C., & Leppink, J. (2018). Cognitive load and learning in the study of multiple documents. *Frontiers in Education*, *3*(59), 1–7. https://doi.org/10.3389/feduc.2018.00059

Cuevas, I., Mateos, M., Martín, M., Luna, M., Martín, A., Solari, M., & Martínez, I. (2016). Collaborative writing of argumentative syntheses from multiple sources: The role of writing beliefs and strategies in addressing controversy. *Journal of Writing Research*, *8*(2), 205–226. https://doi.org/10.17239/jowr-2016.08.02.02

De Smet, M. J. R., Leijten, M., & Van Waes, L. (2018). Exploring the process of reading during writing using eye tracking and keystroke logging. *Written Communication*, *35*(4), 411–447. https://doi.org/10.1177/0741088318788070

Dochy, F., Segers, M., & Buehl, M. M. (1999). The relation between assessment practices and outcomes of studies: The case of research on prior knowledge. *Review of Educational Research*, *69*(2), 145–186. https://doi.org/10.3102/00346543069002145

Elff, D. (2020). *Mclogit: Multinomial logit models, with or without random effects or overdispersion*. R package version 0.8.6.4. https://CRAN.R-project.org/package=mclogit

Escorcia, D., Passerault, J. M., Ros, C., & Pylouster, J. (2017). Profiling writers: Analysis of writing dynamics among college students. *Metacognition and Learning*, *12*(12), 233–237. https://doi.org/10.1007/s11409-016-9166-6

Galbraith, D. (1992). Conditions for discovery through writing. *Instructional Science*, *21*(1–3), 45–72. https://doi.org/10.1007/BF00119655

Galbraith, D. (1999). Writing as a knowledge-constituting process. In M. Torrance & D. Galbraith (Eds.), *Knowing what to write: Conceptual processes in text production* (pp. 139–160). Amsterdam University Press.

Graves, R., Hyland, T., & Samuels, B. M. (2010). Undergraduate writing assignments: An analysis of syllabi at one Canadian college. *Written Communication*, *27*(3), 293–317. https://doi.org/10.1177/0741088310371635

Harring, J. R., & Hodis, F. A. (2016). Mixture modeling: Applications in educational psychology. *Educational Psychologist*, *51*(3–4), 354–367. https://doi.org/10.1080/00461520.2016.1207176

Hartley, J., & Branthwaite, A. (1989). The psychologist as wordsmith: A questionnaire study of the writing strategies of productive British psychologists. *Higher Education*, *18*(4), 423–452. https://doi.org/10.1007/BF00140748

Hayes, J. R., & Flower, L. (1980). Identifying the organization of writing processes. In L. Gregg & E. R. Steinberg (Eds.), *Cognitive processes in writing* (pp. 3–30). Erlbaum.

Hayes, J. R., & Nash, J. G. (1996). A new model of cognition and affect in writing. In C. M. Levy & S. Ransdell (Eds.), *The science of writing: Theories, methods, individual differences and applications* (pp. 29–56). Erlbaum.

Hickendorff, P. A., Edelsbrunner, J., McMullen, M., Schneider, M., & Trezise, K. (2018). Informative tools for characterizing individual differences in learning: Latent class, latent profile, and latent transition analysis. *Learning and Individual Differences*, *66*, 4–15. https://doi.org/10.1016/j.lindif.2017.11.001

Jonsson, A., & Svingby, A. (2007). The use of scoring rubrics: Reliability, validity and educational consequences. *Educational Research Review*, *2*(2), 130–144. https://doi.org/10.1016/j.edurev.2007.05.002

Karlsson, J., van den Broek, P., Helder, A., Hickendorff, M., Koorneef, A., & van Leijenhorst, L. (2018). Profiles of young readers: Evidence from thinking aloud while reading narrative and expository texts. *Learning and Individual Differences*, *67*, 105–116. https://doi.org/10.1016/j.lindif.2018.08.001

Kellogg, R. T. (1987). Effects of topic knowledge on the allocation of processing time and cognitive effort to writing processes. *Memory & Cognition*, *15*(3), 256–266. https://doi.org/10.3758/BF03197724

Kieft, M., Rijlaarsdam, G., Galbraith, D., & van den Bergh, H. (2007). The effects of adapting a writing course to students' writing strategies. *The British Journal of Educational Psychology*, *77*(3), 565–578. https://doi.org/10.1348/096317906X120231

Kim, H. (2020). Profiles of undergraduate student writers: Differences in writing strategy and impacts on text quality. *Learning and Individual Differences*, *78*, 101823. https://doi.org/10.1016/j.lindif.2020.101823

Knospe, Y. (2017). *Writing in a third language. A study of upper-secondary students' texts, writing processes and metacognition* (Doctoral dissertation). Umeå University.

Lavelle, E. (2007). Approaches to writing. In M. Torrance & M. Galbraith (Eds.), *Writing and cognition: Research and applications* (pp. 219–230). Elsevier.

Leijten, M., & Van Waes, L. (2013). Keystroke logging in writing research: Using Inputlog to analyze and visualize writing processes. *Written Communication*, *30*(3), 358–392. https://doi.org/10.1177/0741088313491692

Leijten, M., Van Waes, L., Schrijver, I., Bernolet, S., & Vangehuchten, L. (2017). Hoe schrijven masterstudenten syntheseteksten? Het brongebruik van gevorderde schrijvers in kaart gebracht [How do master's students write synthesis texts? Source use by advanced writers]. *Pedagogische Studiën*, *94*(4), 233–253.

Leijten, M., Van Waes, L., Schrijver, I., Bernolet, S., & Vangehuchten, L. (2019). Mapping master's students' use of external sources in source-based writing in L1 and L2. *Studies in Second Language Acquisition*, *41*(3), 555–582. https://doi.org/10.1017/S0272263119000251

Leijten, M., Van Waes, L., Schriver, K., & Hayes, J. R. (2014). Writing in the workplace: Constructing documents using multiple digital sources. *Journal of Writing Research*, *5*(3), 285–337. https://doi.org/10.17239/jowr-2014.05.03.3

Lenth, R. (2020). Emmeans: Estimated marginal means, aka least-squares means. R package version 1.4.6. https://CRAN.R-project.org/package=emmeans

Lesterhuis, M., Verhavert, S., Coertjens, L., Donche, V., & De Maeyer, S. (2016). Comparative judgment as a promising alternative to score competences. In E. Cano & G. Ion (Eds.), *Innovative practices for higher education assessment and measurement* (pp. 119–138). IGI Global. https://doi.org/10.4018/978-1-5225-0531-0

Levy, M. C., & Ransdell, S. (1996). Writing signatures. In C. M. Levy & S. Ransdell (Eds.), *The science of writing: Theories, methods, individual differences, and applications* (pp. 149–162). Erlbaum.

Lindgren, E., & Sullivan, K. (2019). *Observing writing. Insights from keystroke logging and handwriting.* Brill. https://doi.org/10.1163/9789004392526

List, A., & Alexander, P. A. (2017). Text navigation in multiple source use. *Computers in Human Behavior*, *75*, 364–375. https://doi.org/10.1016/j.chb.2017.05.024

Martínez, I., Mateos, M., Martín, E., & Rijlaarsdam, G. (2015). Learning history by composing synthesis texts: Effects of an instructional programme on learning, reading and writing processes, and text quality. *Journal of Writing Research*, *7*(2), 275–302. https://doi.org/10.17239/jowr-2015.07.02.03

Mateos, M., & Solé, I. (2009). Synthesising information from various texts: A study of procedures and products at different educational levels. *European Journal of Psychology of Education*, *24*(4), 435–451. https://doi.org/10.1007/BF03178760

Mateos, M., Martín, E., Villalón, R., & Luna, M. (2008). Reading and writing to learn in secondary education: Online processing activity and written products in summarizing and synthesizing tasks. *Reading and Writing*, *21*(7), 675–697. https://doi.org/10.1007/s11145-007-9086-6

Mateos, M., Solé, I., Martín, E., Cuevas, I., & Castells, N. (2014). Writing a synthesis from multiple sources as a learning activity. In P. Klein, P. Boscolo, L. Kirkpatrick, & C. Gelati (Eds.), *Writing as a learning activity* (pp. 168–190). Brill. https://doi.org/10.1163/9789004265011_009

Merchie, E., & Van Keer, H. (2014). Using on-line and off-line measures to explore fifth and sixth graders' text-learning strategies and schematizing skills. *Learning and Individual Differences*, *32*, 193–203. https://doi.org/10.1016/j.lindif.2014.03.012

Piolat, A. (1999). Planning and text quality amongst undergraduate students: Findings and questions. In M. Torrance & D. Galbraith (Eds.), *Knowing what to write: Conceptual processes in text production* (pp. 121–138). Amsterdam University Press.

Plakans, L. (2009). The role of reading strategies in integrated L2 writing tasks. *Journal of English for Academic Purposes*, *8*(4), 252–266. https://doi.org/10.1016/j.jeap.2009.05.001

R Core Team. (2019). *A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Raedts, M., Van Steendam, E., De Grez, L., Hendrickx, J., & Masui, C. (2017). The effects of different types of video modelling on undergraduate students' motivation and learning in an academic writing course. *Journal of Writing Research*, *8*(3), 399–435. https://doi.org/10.17239/jowr-2017.08.03.01

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* (pp. 107–126). Erlbaum.

Rijlaarsdam, G., Janssen, T., Braaksma, M., Van Steendam, E., Van den Branden, K., & Verheyden, L. (2014). Learning and instruction in writing. In C. Addison Stone, E. R. Silliman, B. J. Ehren, & G. P. Wallach (Eds.), *Handbook of language and literacy, development and disorders* (pp. 545–566). Guilford.

Rosenberg, J. M. (2019). tidyLPA: An R Package to easily carry out latent profile analysis (LPA) using open-source or commercial software. *Journal of Open Source Software*, *4*(44), 978. https://doi.org/10.21105/joss.00978

Saaty, T. L. (1990). How to make a decision: The analytic hierarchy process. *European Journal of Operational Research*, *48*(1), 9–26. https://doi.org/10.1016/0377-2217(90)90057-I

Scrucca, L., Fop, M., Murphy, T. B., & Raftery, A. E. (2016). Mclust 5: Clustering, classification and density estimation using Gaussian finite mixture models. *The R Journal*, *8*(1), 289–317. https://doi.org/10.32614/RJ-2016-021

Snow, E. L., Allen, L. K., Jacovina, M. E., Crossley, S. A., Crossley, S. A., Perret, C. A., & McNamara, D. S. (2015). Keys to Detecting Writing Flexibility over Time: Entropy and Natural Language Processing. *Journal of Learning Analytics*, *2*(3), 40–63. https://doi.org/10.18608/jla.2015.23.4

Solé, I., Miras, M., Castells, N., Espino, S., & Minguela, M. (2013). Integrating Information: An analysis of the processes involved and the products generated in a written synthesis task. *Written Communication*, *30*(1), 63–90. https://doi.org/10.1177/0741088312466532

Spivey, N., & King, J. (1989). Readers as writers composing from sources. *Reading Research Quarterly*, *24*(1), 7–26. https://doi.org/10.1598/RRQ.24.1.1

Strømsø, H. I., Bråten, I., & Brante, E. W. (2020). Profiles of warm engagement and cold evaluation in multiple-document comprehension. *Reading and Writing*, *33*(9), 2337–2359. https://doi.org/10.1007/s11145-020-10041-5

Tillema, M. (2012). *Writing in a first and second language. Empirical studies on text quality and writing processes* [Doctoral dissertation]. Utrecht University.

Torrance, M., Thomas, G. V., & Robinson, E. J. (1994). The writing strategies of graduate research students in the social sciences. *Higher Education*, *27*(3), 379–392. https://doi.org/10.1007/BF03179901

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*(2), 189–199. https://doi.org/10.1348/000709999157662

Torrance, M., Thomas, G. V., & Robinson, E. J. (2000). Individual differences in undergraduate essay-writing strategies: A Longitudinal study. *Higher Education*, *39*(2), 181–200. https://doi.org/10.1023/A:1003990432398

Van den Bergh, H., & Rijlaarsdam, G. (1996). The dynamics of composing: Modeling writing process data. In C. M. Levy & S. Ransdell (Eds.), *The science of writing* (pp. 207–232). Erlbaum.

Van den Bergh, H., & Rijlaarsdam, G. (1999). The dynamics of idea generation during writing: An online study. In M. Torrance & D. Galbraith (Eds.), *Knowing what to write: Conceptual processes in text production* (pp. 99–120). Amsterdam University Press.

Van den Bergh, H., De Maeyer, S., van Weijen, D., & Tillema, M. (2012). Generalizability of text quality scores. In E. Van Steendam, M. Tillema, G. Rijlaarsdam, & H. van den Bergh (Eds.), *Measuring writing: Recent insights into theory, methodology and practices* (pp. 23–32). Brill.

Van den Bergh, H., Rijlaarsdam, G., & Van Steendam, E. (2016). Writing process theory: A functional dynamic approach. In C. MacArthur, S. Graham & J. Fitzgerald (Eds.), *Handbook of writing research* (2nd ed., pp. 57–71). Guilford Press.

Van der Hoeven, J. (1997). *Children's composing: A study into the relationship between writing processes, cognitive and linguistic skill and text quality*. Rodopi. https://doi.org/10.1163/9789004484689

Van Ockenburg, L., Van Weijen, D., & Rijlaarsdam, G. (2019). Learning to write synthesis texts: A review of intervention studies. *Journal of Writing Research*, *10*(3), 401–428. https://doi.org/10.17239/jowr-2019.10.03.01

Van Waes, L., & Schellens, P. J. (2003). Writing profiles: The effect of the writing mode on pausing and revision patterns of experienced writers. *Journal of Pragmatics*, *35*(6), 829–853. https://doi.org/10.1016/S0378-2166(02)00121-2

Van Weijen, D. (2009). *Writing processes, text quality, and task effects. Empirical studies in first and second language writing* [Doctoral Dissertation]. Utrecht University.

Vandermeulen, N., De Maeyer, S., Van Steendam, E., & Rijlaarsdam, G. (2020). Mapping synthesis writing in various levels of Dutch upper-secondary education. A national baseline study on text quality, writing process and students' perspectives on writing. *Pedagogische Studien*, *97*(3), 187–236.

Vandermeulen, N., Van den Broek, B., Van Steendam, E., & Rijlaarsdam, G. (2020). In search of an effective source use pattern for writing argumentative and informative synthesis texts. *Reading and Writing*, *33*(2), 239–269. https://doi.org/10.1007/s11145-019-09958-3

Vandermeulen, N., Van Steendam, E., De Maeyer, S., & Rijlaarsdam, G. (2022). *Writing process feedback based on keystroke logging and comparison with exemplars: Effects on text quality and process of synthesis texts*. Manuscript submitted for publication.

Verheyden, L. (2010). *Achter de lijn. Vier empirische studies over ontluikende stelvaardigheid* [Behind the line: Four empirical studies about emergent written composition skill] (Unpublished Doctoral Dissertation). KU Leuven.

Villalón, R., Mateos, M., & Cuevas, I. (2015). High school boys' and girls' writing conceptions and writing self-efficacy beliefs: What is their role in writing performance? *Educational Psychology*, *35*(6), 653–674. https://doi.org/10.1080/01443410.2013.836157

Wengelin, A. (2006). Examining pauses in writing: Theory, methods and empirical data. In K. H. Sullivan & E. Lindgren (Eds.), *Computer keystroke logging and writing: Methods and applications* (pp. 107–130). Elsevier.

White, M. J., & Bruning, R. (2005). Implicit writing beliefs and their relation to writing quality. *Contemporary Educational Psychology*, *30*(2), 166–189. https://doi.org/10.1016/j.cedpsych.2004.07.002

(*Appendices follow*)

Appendix A

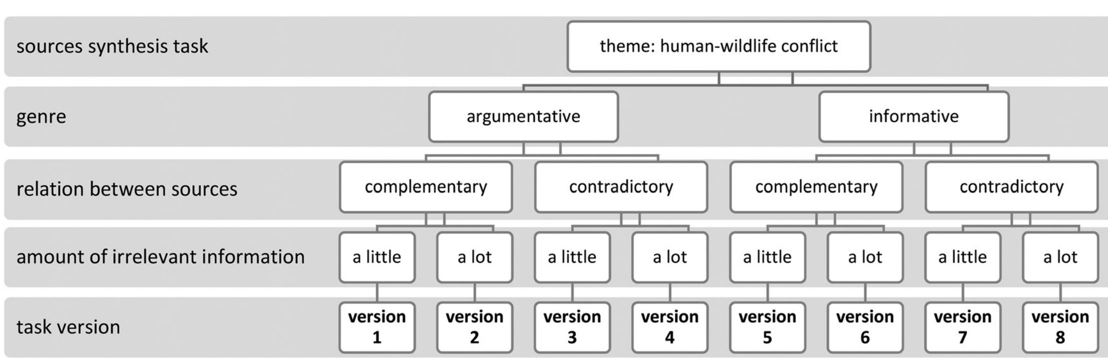
Task Instructions for the Participants

Type of instruction	Clarification
Explanation on what a synthesis text is	A synthesis is a text based on various sources. Your text brings together the information from the different sources. When reading a synthesis you should be able to understand the text without having read the sources.
	How do you write a synthesis?
	- You start by reading the sources
	- You select the information you need, to write a new text about theme X.
	- You bring together the information from the different sources and connect the sources. In this way you integrate the information from the sources into a new independent text.
	- You elaborate your synthesis by writing a text that is understandable for people who have not read the sources.
Explanation on the characteristics of an argumentative/ informative synthesis text	- Informative synthesis: Your text gives a concise and at the same time clear overview of the situation. You describe the situation concerning theme X in a neutral manner, that is, without taking position.
	- Argumentative synthesis: In your text you defend the following point of view: X. You support this point of view with arguments from the source texts.
Instructions on how to deal with the sources	Use the relevant information and use information from all offered source texts.
Instructions on the target audience	Your text has to be understandable to peers who did not read the source texts.
Instructions on style	Use your own words, avoid copying from the sources.
Instructions on text length	Write a text of approximately 350 words
Time indication	You have 50 minutes to read the sources and to write your text. Divide your time between reading and writing. Write the best possible text in this given time.

Note. "Mapping Synthesis Writing in Various Levels of Dutch Upper-Secondary Education: A National Baseline Study on Text Quality, Writing Process and Students' Perspectives on writing," by N. Vandermeulen, S. De Maeyer, E. Van Steendam, & G. Rijlaarsdam, 2020, *Pedagogische Studien*, *97*(3), p. 215. Copyright 2020 by Nina Vandermeulen et al. (LIFT-project). Reprinted with permission.

Appendix B

Visualization of the Task Construction for One Topic



Note. "Mapping Synthesis Writing in Various Levels of Dutch Upper-Secondary Education: A National Baseline Study on Text Quality, Writing Process and Students' Perspectives on writing," by N. Vandermeulen, S. De Maeyer, E. Van Steendam, & G. Rijlaarsdam, 2020, *Pedagogische Studien*, *97*(3), p. 214. Copyright 2020 by Nina Vandermeulen et al. (LIFT-project). Reprinted with permission.

(*Appendices continue*)

Appendix C

Example of a Rating Scale With Benchmark Texts for Synthesis Texts

50

Undangered animals

You might rember the whole internet going crazy a couple of months ago because Cecil the lion got shot dead by an American dentist who took bits of him back to America.

This is a perfect example of trophy hunting. In some country's it's actually legal. The money that's made from it goes straight in the pockets of people who don't have hardly anything to do with protecting animals or nature. Except trophy hunting, there is another from of animal hunting, poaching. Most people think that poachers are just individual people who are not well armed, but the opposite is were poachers are well-organised gangs who have heavy weapons and sometimes even helicopters. All this for certain body parts like a rhinos tusk, which the poachers can sell for loads of money.

When people start living at the edge of nature reserves its obvious that conflicts happen between humans and animals, that's what happened in the Masia Mara national nature reserve. A group of lions killed some cows at night. The farmers knew that the lions would come back to eat more of the bodies, so they sprinkled the carcasses with poison. 3 lions died and the other lions were only just saved by the team of vets. There are a few pretty simple solutions that could stop this happening for example better livestock manaegment or the use of specially trained dogs. This would mean less conflicts, so less dead animals.

To finish I would like to show you a table where you can read how quickly the number of animals are decreasing. You can draw your own conclusions.

year	elephant	black rhino	lion	giraffe	ceetah
1960	1300 000	100 000	200 000	300 000	40 000
2000	600 000	7000	25 000	140 000	20 000
2015	400 000	3050	15 000	80 000	7000

Information
There is a lack of key information from the sources that is necessary for a proper understanding of this text. The key ideas are not clearly or comprehensively described (e.g. it is not clear to the reader what exactly the three threats to wild animals are (trophy hunting, poaching and human/animal conflicts), nor are the solutions sufficiently addressed). The text also contains information that is irrelevant (e.g. dentist who killed Cecil, three lions dead and others rescued by the vet) or incorrect.
Integration
The sources are not integrated into the text. The text has no overarching theme. Content from the sources is presented separately; no clear links are made between the sources (e.g. the table is a source that has simply been 'pasted' into the text without any attempt at integration).
Structure
Both coherence and cohesion are very weak. There is little coherence in terms of content; there is no line of reasoning. As a result, it is not clear what the subject of the text is. The text has not been divided into paragraphs correctly. There is no introduction and no conclusion. The title is too vague and too general.
Language
Many language errors (e.g. spelling). Also incorrect word use/expressions (e.g. the internet going crazy). The text contains sentences that have been copied from the sources word for word. The style is not appropriate. There are also punctuation errors.

75

In this text I am going to give you some information about the situation of endangred African animals.

In Kenya 3 lions from the famous pride Marsch Pride have been found dead. The lions had killed cows belonging to a Masai family at night. In the morning the farmers found the cows bodies and poured poison over them. The lions came back and ate more of the cows, that's how they got poisoned.

It is not the first time, however, that endangered wild animals have been killed by people. There is also such a thing as trophy hunting. This is a legal hunt for pre-selected wild animals, often large animals, the remains of which – like the head, skin, legs or whole body – are taken back to the country where the hunter comes from.

Some travel agencies organise hunting trips, and African governments sell hunting licences to 'fund nature conversation projects'. But the money doesn't go to nature conservation, it disappears immediately into the pockets of corrupt officials or owners of private reserves. People who don't care one bit about how nature is doing.

The poaching of wild animals is a major threat to the survival of many species. Endangered animals are slaughtered by well-organised and heavily armed gangs of poachers. They use heavy weapons to kill the animals and to also defend themselves from the patrolling rangers that protect the African countryside.

It can be seen from the figures that there are less and less wild animals in Africa. The black rhino, for example. In 1960 there were 100,000 and in 2015 there where only 3050. In China and Vietnam people believe that a rhinoceros horn is a powerful medicine against impotence, fever and cancer. That's why lots of rhinos are killed. The body parts are sold on the black market for a lot of money.

After the death of a famous lion (C ecil) many celebrities get involved. Leonardo DiCaprio, Nicole Kidman, and the British princes Harry and William, demand the US to recognize certain animals as officially endangered, then the remains of the trophy hunts aren't allowed to enter the country anymore.

The local population can also do something too. For example, the local population can get involved and actively play a role in tourism and nature conservation. In this way they become more aware of the importance of looking after nature and wildlife.

Information
Key information is missing from the sources in this text. The key ideas are not presented clearly or are missing (e.g. the human/animal conflict is not mentioned, no information on certain solutions such as better livestock management). The key ideas are not described clearly enough (e.g. it is not clear to the reader what trophy hunting, poaching and the human/animal conflict mean exactly). The text contains irrelevant or incorrect information.
Integration
Integration of the sources is weak. The text has no clear overarching theme. An attempt has been made to group information from the sources (e.g. final paragraph brings together some of the possible solutions). In addition, the information is largely presented source by source. Few links are made between the different sources.
Structure
Both coherence and cohesion are weak. There is no clear coherence in terms of content and no clear line of reasoning (e.g. the information on poaching is not grouped together but dotted around throughout the text). The text has not been divided into paragraphs correctly and there is a lack of cohesive devices. There is no real introduction (only one sentence) and no conclusion. There is no title.
Language
Some language errors (both spelling and grammar). The text contains many sentences copied from the sources. The sentences that the author has produced are difficult to read. Style is ok. Punctuation could be better.

100
Average text

Is the drop in the amount of wild animals in Africa a growing problem?

Africa is the poorest continent in the world, this is partly because of the lack of fertile land and also because of the lack of goods to trade. Africa is very poor in figures but compensates for that with its natural habitats, beautiful animals live there which makes the continent as a whole unique. It is not only the african people who know the value of their countryside. There are also Europeans and Americans who want to take advantage of it. All this is at the expense of the animals in Africa. How bad is the situation in Africa?

According to figures, trophy hunting is an ever increasing problem in Africa. Rich europeans and Americans buy hunting licences from African countries and can therefore hunt large unique animals legally.

Pouching is also one of the reasons why the number of animals in Africa has decreased so much in the last few years. In Asia, there is a high demand for certain body parts of endangered species. Large criminal organisations capitalise on this poach elephants, rhinoceroses and lions with heavy artillery.

In addition to trophy hunting, the threat to wildlife in Africa is not only caused by rich Westerners. The threat posed by the wild animals themselves to the cattle owned by the indiginous population is also a reason why the situation is getting worse. Indigenous people lose crops and animals due to elephants trampling around and hungry lions, sometimes they try to take back control and shoot a wild animal.

The African Wildlife foundation says in a scientific report that a sustainable solution to Africa's growing problems is to make agreements with NGOs and governments, as well as better training for indigenous people to ensure that there are fewer losses for them. Finally, the governments of the countries have to put more money into large projects.

In short, the continent of Africa is suffering from a big decrease in animal numbers. This is down to trophy hunters, poachers and the destruction caused by the animals, which makes the indigenous population take action. According to the African Wildlife Foundation, the solution is to improve communication between governments and NGO's, governments have to put more money into projects against poachers and for nature and finally, the local population must be better trained to deal with this situation.

Information
The text contains all of the key information from the sources. The key ideas are described briefly. The information does remain limited, however, and could be clearer (e.g. the three threats to wildlife are mentioned but could be defined more clearly; the information on solutions is incomplete and vague). The information is mostly relevant and correct. Elements that are less relevant or not completely correct do not disturb our understanding (examples of less relevant info: scientific report by the African Wildlife Foundation, info in introduction on poverty in Africa; examples of incorrect info: "According to figures, trophy hunting is an ever increasing problem": this is not mentioned in the sources).
Integration
An attempt has been made to integrate the sources. The text has an overarching theme, but the sources are not fully integrated into a new structure. Instead, the text is a list of summaries of the different sources. There is one source (difficult to integrate) that is not integrated into the text (table with figures).
Structure
Coherence in terms of content is quite good, but there is room for improvement. There is a line of reasoning that makes the information in the text easy to follow, but the transitions between the paragraphs are sometimes quite abrupt. Cohesive devices are used (e.g. also, in addition to, finally, in short). The text also has an attempt at an introduction and conclusion. However, the focus on the theme of the text could be better in the introduction. The conclusion is actually a repetition of what was said in the previous paragraph. The title doesn't completely cover what the text is about.
Language
The text contains a number of language errors (e.g. use of capital letters), though they don't disturb the reader too much. Sentence structure and word choice are ok, but could be better. Style is ok. Punctuation is mostly correct.

125

Wild animals are coming under threat in Africa

As we know, many animals in Africa have been labelled as endangered species. It is therefore very important to protect these animals. If some animals died out an entire ecological area could be endangered because of the fact that different animal species are often very much in tune with each other. The sudden extinction of an animal would be disastrous.

Wild animals don't have it easy in Africa. After all, they are threatened by a number of key factors. First of all, of course, the best known, the poaching of wild animals by poachers. The poachers murder animals for their body parts, for example an elephant for its tusks. They can then sell it for lots of money. There are 'Rangers' in the area who try to protect the area, but the poachers are often heavily armed. This is to kill the animals of course, but also to be able to defend themselves.

Another activity is what's known as 'Trophy hunting'. This is the legal hunting of pre-selected animals, the remains of which are then taken back to the hunter's country as a sort of trophy. The government of a country sells the hunting licences supposedly to be able to fund nature conservation, but the money often ends up in the pockets of corrupt officials.

Finally, there is also the conlict with the local population. Because many of the wild animals are located on the edges or outside of protected reserves, they pose a threat to the local popylation. For example, they destroy crops and kill cattle. The local population often takes matters into its own hands when it comes to this threat by killing the animals themselves. There are some solutions to this problem, such as better lifestock management, specially trained dogs and getting the population involved. And they work, because the farmers have already killed 80% less wild animals since the dogs started been used.

As you can see, there are many factors that threaten wild animals. Some, such as the conflict with the local population, already have solutions. The other factors, such as poaching and trophy hnting, have less easy solutions, but they will also have to be solved. This is needed to safeguard natural habitats in Africa.

Information
The text contains all of the key information from the sources. Most of the key ideas are concisely and clearly defined (e.g. threats to wildlife and solutions to the human/animal conflicts). Some ideas are not addressed (e.g. solution to trophy hunting) or not sufficiently clear (e.g. solution to poaching) in the text. There is no irrelevant or incorrect information.
Integration
The text has an overarching theme. Information from the sources is integrated into a text which has its own structure and links the sources to each other. There is one source (difficult to integrate) that is not integrated into the text (table with figures).
Structure
Coherence in terms of content is good – the text has a clear line of reasoning. Cohesion is also good: correct paragraph structure and good use of cohesive devices. The introduction sets the tone, but could be more specific, i.e. a better focus on the content that follows (e.g. theme only becomes clear in the first sentence of the second paragraph). Nice conclusion. The title is good.
Language
Some language errors, a bit of sloppiness. Use of words is good, but could be more varied. Style is ok. Punctuation is mostly correct.

150

How do you protect endangered species?
An ivory figurine, a panther skin rug and a bison horn. What do these objects have in common? They're all objects made from the body parts of endangered species. These animals are sometimes killed by poachers, local farmers and 'trophy hunters'.

Trophy hunting is a phenomenon where hunters from other countries pay to be allowed to shoot endangered animals legally. Afterwards, the hunters can take 'their' animal, or parts of it home as a 'trophy'. Trophy hunting is a simple way for the governments of poor African countries to fund nature conservation. However, in reality the money disappears into the pockets of corrupt officials.

Not only trophy hunting, but also poaching poses a threat to animal species in the African savannahs. Poachers kill these endangered organisms to sell the different parts of the animals on the black market for large amounts of money. These poachers operate in well-organised gangs with helicopters and night vision goggles to find the animals and evade the rangers that protect them.

Furthermore, the local population shoot a significant number of wild animals and in some cases the animals are even poisoned. Wild animals are often butchered because they trample the farmer's fields and devour their crops, or if they have eaten some of the farmers' cattle because the animals are outside their reserve.

As a result of poaching and trophy hunting as well as local farmers, the population of some endangered species has dropped dramatically. For example, in 1960 there were 1 300 000 elephants in Africa, and now there are only 400 000. One way to prevent a further decrease in the number of wild animals is to involve the local population in nature conservation. This makes them aware of the importance of wild animals, as well as providing them with work as rangers. In this way, they not only shoot fewer animals themselves but also prevent poachers from shooting wild animals. In addition specially trained dogs can be introduced among the farmers to protect their cattle and prevent them from being eaten. A solution for trophy hunting is to prohibit the import of endangered species and their body parts to Western countries, so that trophy hunting is no longer attractive.

Information
The text contains all of the key information from the sources. Moreover, the key ideas are fully and clearly defined (e.g. the three threats to wildlife and the solution to each threat). The information is relevant and correct.
Integration
Good integration of resources. The text has an overarching theme and its own structure, in which the sources are linked to each other (e.g. the figures from the table are not presented in a separate paragraph but interwoven with information from the other sources). The text contains information from every source.
Structure
The text has a nice structure. The content is presented coherently and there is a clear line of reasoning. The formal coherence is also clear. Paragraphs follow each other logically and the structure is well supported by cohesive devices (e.g. not only... but also, furthermore, for example, in addition). The text has a good introduction that acquaints the reader with the theme. The conclusion could be better (e.g. the text ends with the possible solutions – a real conclusion is missing). Title is ok. An indication of location (e.g. Africa) would make it more complete.
Language
Few language errors; the errors that are there do not disturb our reading of the text. Smoothly written text, good choice of words, good style. Written in the author's own words. Punctuation is mostly correct.

Note. "Mapping Synthesis Writing in Various Levels of Dutch Upper-Secondary Education: A National Baseline Study on Text Quality, Writing Process and Students' Perspectives on writing," by N. Vandermeulen, S. De Maeyer, E. Van Steendam, & G. Rijlaarsdam, 2020, *Pedagogische Studien*, *97*(3), p. 217. Copyright 2020 by Nina Vandermeulen et al. (LIFT-project). Adapted with permission. Alternative file formats for Appendix C can be found in the Supplemental Materials.

(*Appendices continue*)

Appendix D

Statistics

Table D1

Correlations Between Inputlog Writing Behavioral Measures

Measure	Time in sources Interval 1	Time in sources Interval 3	Switches between sources Interval 1	Switches between sources Interval 2	Switches between sources and text Interval 2	Time writing Interval 1	Time writing Interval 3	Speed (strokes per minute) Interval 1
Time in sources Interval 1	1.00							
Time in sources Interval 3	0.24	1.00						
Switches between sources Interval 1	0.27	0.09	1.00					
Switches between sources Interval 2	0.08	0.15	0.32	1.00				
Switches between sources and text Interval 2	0.15	0.28	0.27	0.34	1.00			
Time writing Interval 1	−0.84	−0.22	−0.21	−0.03	−0.09	1.00		
Time writing Interval 3	−0.05	−0.41	−0.03	−0.08	−0.11	0.12	1.00	
Speed Interval 1	−0.61	−0.22	−0.08	0.04	0.10	0.69	0.06	1.00
Speed Interval 2	0.02	−0.23	0.06	−0.08	−0.05	0.09	0.23	0.36

Table D2

Fit Indices Used for the Analytic Hierarchy Process to Select the Optimum Number of Profiles (Configurations) Based on Finite Mixture Models Varying in Number of Latent Profiles From 1 to 8[a]

N Configurations	AIC	AWE	BIC	CLC	KIC	C-RIV
1	46,027.31	46,912.26	46,335.78	45,921.31	46,084.31	0.12,189
2	44,813.22	46,602.23[b]	45,435.88	44,596.52	44,925.22	0.12,465
3	44,411.91	47,104.32	45,348.75	44,085.20	44,578.91	0.12,524
4	44,031.97	47,627.64	45,283.00[b]	43,595.35	44,253.97	0.12,581[b]
5	43,890.84	48,389.94	45,456.05	43,344.16	44,167.84	0.12,568
6	43,686.76	49,089.18	45,566.16	43,030.14	44,018.76	0.12,576
7	43,585.93	49,891.67	45,779.51	42,819.36	43,972.93	0.12,556
8	43,468.30[b]	50,677.41	45,976.07	42,591.73[b]	43,910.30[b]	0.12,541

[a] These models assume varying variances and covariances for the process measures in each configuration. [b] Indicates the best fitting model according to this information criterion.

Table D3

Estimated Marginal Means for the 9 Inputlog Writing Behavioral Measures (as z-Scores) Per Latent Profile Based on Mixed Effects Models for Each Variable and Significant Contrasts ($p < .05$) Between Process Configurations Based on Pairwise Comparisons

Measure	Configuration 1	Configuration 2	Configuration 3	Configuration 4	Sig. Contrasts
Time in sources Interval 1	−0.184	−0.463	−0.089	0.407	1 > 2; 1 < 4; 2 < 3; 2 < 4; 3 < 4
Time in sources Interval 3	−0.719	0.354	−0.238	0.377	1 < 2; 1 < 3; 1 < 4; 2 > 3; 3 < 4
Switches between sources Interval 1	−0.257	−0.474	0.348	0.475	1 > 2; 1 < 3; 1 < 4; 2 < 3; 2 < 4; 3 < 4
Switches between sources Interval 2	−0.016	−0.008	0.288	0.414	1 < 3; 1 < 4; 2 < 3; 2 < 4; 3 < 4
Switches between sources and text Interval 2	−0.302	−0.329	0.357	0.443	1 < 3; 1 < 4; 2 < 3; 2 < 4; 3 < 4
Time writing Interval 1	0.419	0.168	−0.148	−0.187	1 > 2; 1 > 3; 1 > 4; 2 > 3; 2 > 4
Time writing Interval 3	0.213	−0.100	−0.209	0.131	1 > 2; 1 > 3; 2 < 4; 3 < 4
Speed Interval 1	0.274	−0.022	−0.022	−0.153	1 > 2; 1 > 3; 1 > 4; 2 > 4; 3 > 4
Speed Interval 2	0.363	0.072	−0.276	−0.013	1 > 2; 1 > 3; 1 > 4; 2 > 3; 3 < 4

(*Appendices continue*)

Table D4

Model Fit and Comparisons for Multinomial Models Predicting "Writing Process Configurations"

Model	Deviance	Comparison	$\Delta\chi^2$	Δdf	p
0 Null Model	4,789.0				
1 + Main effect genre	4,714.8	0 vs 1	74.16	3	<.001
2b + Main effects task experiences (Model 2b)	4,621.3	1 vs 2	14.907	12	.247
3b + Interaction effects all task experiences × Genre	4,606.4	2 vs 3	29.99	12	.003
4b + Interaction Effects Topic Prior Knowledge × Genre & Topic interest × Genre	4,576.5	3 vs 4	16.15	12	.185

Table D5

Parameter Estimates (Est.), Standard Errors (SE), z Values, and p Values for the Estimates of the Fixed Part of Multinomial Model 1 Predicting "Writing Process Configurations"

Measure	Est.	*SE*	*z* value	*p* value
Configuration 2 vs. 1				
Intercept[a]	0.335	0.432	0.776	.438
Genre	−0.519	0.656	−0.791	.429
Configuration 3 vs. 1				
Intercept[a]	0.507	0.398	1.275	.202
Genre	0.757	0.583	1.297	.195
Configuration 4 vs. 1				
Intercept[a]	0.533	0.402	1.327	.185
Genre	1.281	0.574	2.231	.026

Note. Configuration 1 = fast text production; Configuration 2 = last-minute source inspection; Configuration 3 = quick source use, and Configuration 4 = source use.
[a] Reference category: Configuration 1.

Table D6

Parameter Estimates (Est.), Standard Errors (SE), t Values, and p Value for the Estimates of the Fixed Part of the Multinomial Model 1 Predicting "Text Quality"

Measure	Est.	*SE*	*z* value	*p* value
Intercept[a]	92.462	1.347	68.619	<.001
Configuration 2	−3.227	1.196	−2.698	.007
Configuration 3	−1.471	1.118	−1.316	.188
Configuration 4	−3.195	1.100	−2.906	.004

Note. Configuration 1 = fast text production; Configuration 2 = last-minute source inspection; Configuration 3 = quick source use, and Configuration 4 = Source use.
[a] Reference category: Configuration 1.

Table D7

Parameter Estimates (Est.), Standard Errors (SE), t Values, and p Value of General Linear Analysis With Gamma Distribution Including Number of Distinct Configurations Used by a Writer on Variability (SD) of Text Quality

Measure	Est.	*SE*	*z* value	*p* value
Intercept[a]	12.809	0.598	21.388	<.001
Two configurations	−0.293	0.704	−0.417	.677
Three configurations	0.755	0.828	0.911	.363
Four configurations	−0.390	2.024	−0.192	.847

[a] Reference category: One configuration.

Table D8

Parameter Estimates (Est.), Standard Errors (SE), z Values, and p Values for the Estimates of the Fixed Part of Multinomial Model 4b Predicting "Writing Process Configurations"

Measure	Est.	*SE*	*z* value	*p* value
Configuration 2 vs. 1				
Intercept[a]	0.335	0.432	0.776	.438
Genre	−0.519	0.656	−0.791	.429
Task interest	−0.246	0.117	−2.109	.035
Task topic knowledge	0.088	0.116	0.755	.450
Task Interest × Genre	0.225	0.179	1.257	.209
Task Topic Knowledge × Genre	−0.028	0.172	−0.163	.871
Configuration 3 vs. 1				
Intercept[a]	0.507	0.398	1.275	.202
Genre	0.757	0.583	1.297	.195
Task interest	−0.216	0.108	−1.991	.047
Task topic knowledge	0.095	0.106	0.896	.370
Task Interest × Genre	0.262	0.161	1.629	.103
Task Topic Knowledge × Genre	−0.345	0.156	−2.210	.027
Configuration 4 vs. 1				
Intercept[a]	0.533	0.402	1.327	.185
Genre	1.281	0.574	2.231	.026
Task interest	−0.044	0.108	−0.402	.687
Task topic knowledge	−0.096	0.107	−0.897	.370
Task Interest × Genre	−0.032	0.157	−0.202	.840
Task Topic Knowledge × Genre	−0.087	0.154	−0.568	.570

Note. Configuration 1 = fast text production; Configuration 2 = last-minute source inspection; Configuration 3 = quick source use, and Configuration 4 = source use.
[a] Reference category: Configuration 1.

(*Appendices continue*)

Appendix E

Selection of Writing Behavior Variables for Profile Analysis

Variables Selection

The selection of behavioral variables was based on two preliminary studies on samples of the dataset, published as Vandermeulen, De Maeyer, et al. (2020) and Vandermeulen, Van den Broek, et al. (2020). These reports were based on a subsample of the texts, 150 texts from both genres (Informative, Argumentative).

We selected five variables from the set of options, based on two preliminary studies, in which relations were shown between processes and text quality, and two additional criteria (see Table E1):

1. Scores must be available on writing process interval level, as former studies have shown that it is not so much the frequency of activities that predicts text quality, but the distribution of these activities across the process, that is, the frequency per interval, which could result in positive correlations in one interval, and negative correlations in another interval. Since Breetvelt et al. (1994) showed via think-aloud studies the dynamic relation between cognitive behaviors and text quality, many other studies confirmed this approach (Rijlaarsdam et al., 2014). This criterion led to the decision to refrain from revision scores, as Inputlog did not automatically generate scores for process intervals. Furthermore, it turned out that for source-based writing, revision scores were invalid because copy and paste actions from sources into the text-so-far were not well-registered.

2. Variables selected must have a communicative value to students as the study's larger context is to provide students with process feedback. For that reason, we needed clear, unambivalent and straight-forward indices. Therefore, we refrained from including pause variables in the current study, while the interpretation of this behavior is multi-interpretable, and partly covered by the active counterparts, which are source activities and production activities. From the production variables, we selected a fluency or speed variable, Keystrokes per minute, and a time variable, actual writing time or Proportion of active writing time, as a counterpart of actual time in Sources (Proportion time in the sources).

These choices resulted in three types of variables with five specific variables:

1. Time: Allocation of time, to Sources or to Text production, as proportions of the total process time cf. Proportion time in the sources and Proportion of active writing time subsequently referred to as Time in sources and Time writing.

2. Actions: the switches from source to source, and the extent to which one mediates from sources to synthesis text. These correspond to Transitions (per min) between the sources and Transitions (per min.) between the sources and the synthesis respectively, referred to as Switches between sources and Switches between sources and text.

3. Speed: speed of text production which corresponds to Strokes per minute, subsequently Speed.

Validation of the Selected Variables in the Total Sample

With these types of variables we analyzed the relation with text quality, following the procedures we reported in Vandermeulen, De Maeyer, et al. (2020). We ran analyses for the source variables and the production variables separately, but completely in parallel, and included in all cases the scores for the three intervals.

1. We compared three models for each variable: a null model (random components subjects, task, and school, and no predictors, Genre to control for a genre effect), Model 1 (added linear relations between the process scores and text quality), and Model 2 (added curvilinear relations) (see Table E2).

Source Variables

a. Model 2 fitted best: curvilinear relations between the frequency of processes and text quality.

b. Four out of nine variables did not contribute to the prediction: (Proportion of) Time in (the) sources (Interval 2); (Number of) Switches between the sources (Interval 3), (Number of) Switches between the sources and the synthesis text (Intervals 1 and 3).

Production Variables

a. Model 2 fitted best: curvilinear relations between the frequency of processes and text quality.

b. Three variables contributed to the prediction: Time writing (linear and curvilinear (Interval 1); Speed (Intervals 1 and 2, both linear).

(*Appendices continue*)

Table E1

Writing Behavioral Variables Included in Preliminary Studies (Vandermeulen, Van Den Broek, et al., 2020, on Source Variables, Vandermeulen, De Maeyer, et al., 2020, on Production Variables) and the Current Study

Variable	Whole process	Intervals	Current study
Source behavior variables			
Proportion time in sources	✓	✓	✓
Transitions between sources	✓	✓	✓
Transitions synthesis–sources	✓	✓	✓
General writing behavior measures			
Process time in minutes	✓		
Time spent writing vs. pausing (proportion)	✓		
Total keystrokes produced in phase (proportion)		✓	
Proportion active writing time			✓
Fluency			
Keystrokes per minute	✓	✓	✓
Number of P-bursts per minute	✓		
Length of P-bursts in seconds	✓		
Length of P-bursts in characters	✓		
Pausing			
Mean pause time	✓	✓	
Percentage of total pauses occurring in phase		✓	
Revision			
Percentage produced text in final synthesis	✓		

Table E2

Model Fit and Comparisons

Models			Comparison			
Model	χ^2	*df*	Models	χ^2_{change}	df_{change}	*p*
Source variables						
0. Null model	−9,042.5	6				
1. Linear effect	−9,021.6	15	0 vs. 1	41.934	9	<.001
2. Curvilinear effects	−9,009.0	24	1 vs. 2	25.016	9	.003
2A. Selected variables	−9,012.5	16	2 vs. 2A	6.808	8	.557
Production variables						
0. Null model	−9,042.5	6				
1. Linear effect	−9,021.9	12	0 vs. 1	41.173	6	<.001
2. Curvilinear effects	−9,004.2	18	1 vs. 2	35.437	6	<.001
2A. Selected variables	−9,011.5	10	2 vs. 2A	14.600	8	.067

Appendix F

Descriptives of Task Experiences Questionnaire

Task experience questions	*M*	*SD*
I knew a lot about the subject of the text before writing it (topic knowledge)	2.70	1.02
I find the subject of the text interesting (topic interest)	3.39	1.01
I found the writing task difficult (experienced difficulty)	2.58	0.89
I put a lot of effort into the writing task (effort)	3.50	0.74

Received May 4, 2021
Revision received March 23, 2022
Accepted April 1, 2022 ■