University of Groningen

Exploring the relationship of working memory to the temporal distribution of pausing and revision behaviors during L2 writing

Révész, Andrea; Michel, Marije; Lee, Minjin

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

[Link to publication in University of Groningen/UMCG research database](#)

Exploring the relationship of working memory to the temporal distribution of pausing and revision behaviors during L2 writing

Andrea Révész[1]* (iD), Marije Michel[2] (iD) and Minjin Lee[3] (iD)

[1]University College London, London, UK; [2]University of Groningen, Groningen, Netherlands; [3]Yonsei University, Seoul, Republic of Korea
*Corresponding author. Email: a.revesz@ucl.ac.uk

Abstract

This study examined the extent to which L2 writers with varied working memory display differential pausing and revision behaviors at different periods during writing. The participants were 30 advanced Chinese L2 users of English, who wrote an argumentative essay. While composing, participants' keystrokes and eye-gaze movements were recorded to capture their pausing, revision, and eye-gaze behaviors. The working memory battery included tests of phonological and visual short-term memory and executive functions. We divided the writing process into five equal periods. The results revealed that participants' pausing and revision patterns were consistent with previous findings that planning, linguistic encoding, and monitoring processes dominate the initial, middle, and later composing periods, respectively. Various working memory components had differential effects on pausing depending on period, largely reflecting the predictions of Kellogg's (1996, 2001) model. However, we identified no differences in the temporal distribution of revision behaviors contingent on working memory.

Introduction

During the past two decades, second language (L2) researchers have shown a growing interest in investigating the processes in which L2 writers engage, with much of the research focusing on directly observable features of the writing process such as pausing and revision behaviors. This line of research has identified several factors that may influence writing behaviors including proficiency (Barkaoui, 2019, Gánem-Gutiérrez & Gilmore, 2018; Lee, 2019; Lu, 2022; Révész et al., 2022), typing skill (Barkaoui, 2016), task type/genre (Barkaoui, 2016; Lee, 2019; Michel et al., 2020; Thorson, 2000), task complexity (Lu, 2022; Révész, Kourtali, et al., 2017), and individual differences (Révész, Michel, et al., 2017). Few studies, however, have considered how the time course of writing may affect writing behaviors (see, however, Lu & Révész, 2021; Révész et al.,

2022; Vallejos, 2020), and it remains unknown how individual differences in working memory (WM) may moderate the temporal distribution of writing processes.

From a theoretical perspective, is it is important to understand the influence of WM on pausing and revision behaviors at various periods during the course of writing. Several theoretical models of writing (e.g., Hayes, 1996; Kellogg, 1996) make explicit predictions about the involvement of WM in cognitive writing processes such as planning, linguistic encoding, monitoring, and transcription, which are expected to play differential roles at different points in the writing process (Rijlaarsdam & Van Den Bergh, 1996). Given that cognitive writing processes have been linked to various pausing and revision patterns (Baaijen et al., 2012; Révész et al., 2019), studying the effects of WM on writing behaviors as a function of writing period will help test these theoretical frameworks and their predictions (see below for details). In general, investigating how WM interacts with writing behaviors during the time course of writing will yield useful insights for building models of writing, as the findings will assist in inferring what cognitive processes writers are engaged in throughout the composing process, which is difficult to observe directly (DeKeyser, 2012).

It is also desirable from an applied point of view to gain a deeper understanding of the role of WM at different periods during L2 writing. For example, if L2 writers with low WM are found to face enhanced difficulty at certain writing periods, pedagogical interventions can be designed to help them develop strategies to tackle the challenges they face at various points in the writing process. Information about the effects of WM on composing processes could also inform the development of assistive technologies and guidelines to accommodate L2 writers with lower WM in high-stakes as well as classroom-assessment settings (Kormos, 2021; Michel et al., 2019; see also Granena's and Kormos's contributions to this special issue).

Against this background, our goal in this study was twofold. First, we intended to expand on previous research by examining how pausing and revision behaviors may vary according to writing period (beginning, mid, end periods). Second, we wanted to launch an investigation into the extent to which individual differences in WM may influence the time distribution of pausing and revision behaviors. In the sections to follow, we will first review the theoretical foundations of this research, followed by an overview of previous empirical work related to the study.

Theoretical background

Cognitive models of L1 and L2 writing (Hayes, 1996; Kellogg, 1996) generally see writing as entailing four subprocesses. Planning involves higher order writing operations such as goal setting and retrieving ideas from long-term memory or the task input and organizing these ideas into a coherent plan. During the course of translation or linguistic encoding, writers turn the content planned into linguistic form through lexical retrieval, syntactic encoding, and use of cohesive devices. During execution, writers employ motor movements to create a typed or hand-written text. Finally, monitoring entails controlling the whole process and rereading and editing to check whether the evolving text expresses the writer's intended content. These writing processes are presumed to work in parallel in a cyclical manner. Increasingly, writing is also considered to be a dynamic process in a broader sense—that is, writers are thought to be involved in different writing processes to a differential degree at various points of writing (e.g., beginning, middle, end periods). The temporal distribution of writing processes is assumed to mirror changes in how the writer perceives the task as the writing process proceeds (Khuder & Harwood, 2015; Nicolás-Conesa et al., 2014;

Rijlaarsdam & Van Den Bergh, 1996). In other words, the cognitive activities of writers are expected to vary as a function of the altering task environment.

Some L1 writing models also posit a pivotal role for WM in the writing process. Hayes' (1996) and Kellogg's (1996, 2001) influential models of writing draw on Baddeley's (1986) multicomponent WM framework. This model describes WM as a system composed of a central executive and two domain-specific subsystems, a phonological loop and a visual-spatial sketchpad. A later version of the model (Baddeley, 2000) posits a third subsystem called episodic buffer. The central executive oversees complex cognitive operations including dividing, focusing, and switching attention; activating and suppressing processing routines; and controlling the flow of information from the two subsystems and long-term memory. The phonological loop is involved in temporarily storing and manipulating acoustic and verbal information, whereas the visual-spatial sketchpad is responsible for storing and processing spatial and visual information. The episodic buffer integrates information from the other two slave systems and long-term memory to create multimodal representations (e.g., a story). All these components of WM are presumed to be limited in capacity.

Hayes (1996) argues that all writing operations rely on WM and all nonautomatic writing activities take place in WM. Going further, Kellogg (1996, 2001) makes explicit predictions about the involvement of specific WM components in writing processes. In Kellogg's view, the central executive is implicated in all writing subprocesses, the only exception being execution in cases where writers possess automatic typing or handwriting skills. The phonological loop is called on when writers engage in translation or rereading previously produced text during monitoring, as these processes entail the processing of verbal material. The visual-spatial sketchpad is required during planning content and organization as well as editing as part of the monitoring process. When writers plan, they often generate prelinguistic ideas that frequently involve images, and when they edit, they need to rely on visual and spatial information to organize their text.

Following Kellogg's (1996, 2001) predictions, we would expect that different components of WM will be implicated to various degrees at different periods during writing, as writers are expected to engage in different writing subprocesses to a varied extent throughout the composing process. To inform our subsequent discussion of how various WM components may influence pausing and revision behaviors across writing periods, we now turn to a discussion of what cognitive writing processes may underlie pausing and revision behaviors, followed by a review of previous research investigating the temporal distribution of writing subprocesses and behaviors.

Pausing and revision behaviors and associated cognitive processes

Pausing during writing, defined as a lack of handwriting or typing, has been associated with various physical (e.g., issues with motor movement), sociopsychological (e.g., mind wandering), and cognitive writing processes (e.g., planning, linguistic encoding, and reading the evolving text; Alves et al., 2007). Although it is difficult to identify the precise processes underlying pausing, previous L2 research suggests (e.g., Chukharev-Hudilainen et al., 2019; Révész, Kourtali, et al., 2017; Révész et al., 2019, Spelman Miller, 2000) that, depending on the textual location at which they occur, pauses are more or less likely to reflect certain underlying processes. Specifically, there appears to be a greater likelihood that pausing at higher level textual units (e.g., between clauses and sentences) relates to the writers' planning of content and organization. On the other hand, it seems more probable that pauses at lower textual units (e.g., within and

between words) are associated with linguistic encoding processes, including lexical retrieval and morphological encoding. Therefore, following Kellogg's predictions about how aspects of WM link to cognitive writing processes, we would expect that the central executive relates to pausing behaviors at all textual locations, whereas the phonological loop and visual spatial sketchpad have stronger relationships to pauses at lower and higher textual units respectively.

Turning to revision, the process entails reading, assessing, and visibly altering one's evolving text (external revision) and changing ideas that might have been planned and/or translated within the writer's head before the text is physically altered (internal revision) (Lindgren & Sullivan, 2006; Stevenson et al., 2006). External revisions, the focus of the present study, may occur at the point of inscription (precontextual revision) or away from it (contextual revision). Another way to categorize revisions is in terms of the size of the textual unit that is being revised, whether it involves changing a lower level (e.g., a character or word) or a higher level (e.g., a clause or sentence) unit. Parallel to research on pausing, previous L2 research found that the likelihood of writers engaging in various cognitive writing processes varies according to the level of revision they make. Although L2 writers were found to focus predominantly on linguistic issues regardless of level of revision, planning-related problems were more likely to be addressed through higher level revision, both when researchers considered only precontextual revisions (Révész, Kourtali, et al., 2017) and when they took account of contextual as well as precontextual revisions (Révész, Michel, et al., 2017). Thus, based on Kellogg's predictions regarding the role of WM in writing processes, we would anticipate that the central executive and phonological loop are implicated in all levels of revision, whereas the visual spatial sketchpad is linked more strongly to revision of larger textual units.

With a view toward understanding how the putative links outlined here between WM components and pausing and revision behaviors may vary throughout the writing process, we continue with an overview of prior research examining the time distribution of cognitive writing processes and associated behaviors.

Writing behaviors and the time course of writing

Primarily inspired by the work of Rijlaarsdam and Van den Bergh (1996), the last decade has seen an increasing interest in the temporal distribution of writing behaviors and associated cognitive processes. Earlier studies on the temporal distribution of writing activities have mainly used verbal protocols such as think-aloud and stimulated recall procedures. Manchón, Roca de Larios, and their colleagues (Manchón et al., 2009; Manchón & Roca de Larios, 2007; Roca de Larios et al., 2008) were among the first to explore how cognitive writing activities changed across the time course of writing using think-aloud protocols. The researchers divided the participants' total writing time into three equal periods and compared the type and frequency of cognitive activities in which participants engaged at each interval. Planning, formulation (i.e., linguistic encoding), and revision emerged as the three main writing activities in the think-alouds, taking up approximately 90% of participants' total writing time. Participants produced planning-related comments more frequently when describing their thoughts in the initial as compared with later periods, formulation-related comments reached their peak in the second period, and the number of revision-related comments increased gradually across the three periods.

Subsequent work using verbal protocols has confirmed these trends for planning and linguistic encoding processes, with planning dominating earlier periods and linguistic encoding occurring more frequently during the middle periods of writing (Barkaoui, 2015; Khuder & Harwood, 2015; Michel et al., 2020; Roca de Larios et al., 2008; Tillema, 2012; Van Weijen, 2009). For revision and rereading, the patterns observed are less consistent. In some studies, revision was found to increase across periods (e.g., Barkaoui, 2015; Roca de Larios et al., 2008), whereas in others the distribution of revision was more balanced over time (e.g., Tillema, 2012).

More recently, researchers have also begun to use keystroke-logging software, alone or in combination with other techniques, to study the temporal distribution of pausing and revision behaviors during the course of writing. For instance, Xu and Qi (2017) studied the pausing behaviors of 30 less skilled and 29 more skilled L2 English writers across five periods of argumentative writing. Less-skilled writers paused most frequently in Period 4, and pauses were longer in period 2 as compared with Periods 1, 2, and 4. More-skilled writers, on the other hand, showed less frequent pausing in period 1 than in later periods, but their length of pauses was greater in Period 1 than in Periods 2 to 4. These results were partially confirmed by Barkaoui (2019), who investigated the pausing patterns of 68 English L2 writers during an independent and an integrated writing task. Parallel to Xu and Qi's (2017) findings for skilled writers, Barkaoui observed lower frequency but greater length of pausing during the initial than the middle and end periods of writing across both task types.

To get more complete insights into pausing as a function of time, Michel et al. (2020) triangulated stimulated recall, keystroke-logging, and eye-tracking data to examine, among other things, pausing behaviors and associated cognitive processes. The participants were 60 L2 users, whose composing processes were studied during five periods of two integrated and two independent writing tasks. In the initial period of the independent task, the researchers witnessed slower speed of writing, fewer pauses, and shorter and fewer fixations on the writing window than in the middle periods. Participants wrote faster during Periods 3 and 4 than in the last period and displayed fewer saccades in Period 4 as compared with Period 1. The stimulated recall data, which were elicited to describe participants' thought processes during pauses, included less reference to planning and translation as time progressed, whereas monitoring-related comments increased over time. The researchers interpreted these patterns as an indication that participants engaged in planning in the initial periods, followed by a focus on text production in the middle periods, and monitoring in the final period. During the integrated task, writers showed more dynamic and diverse behaviors and cognitive processes across various periods of writing. One limitation of this study was that the eye-tracking measures were obtained for the entire writing window and for the writing process as a whole rather than for smaller textual units and for the duration of individual pauses. This inevitably resulted in relatively coarse eye-gaze measurements.

Moving onto revision, Barkaoui's (2016) study was one of the first to explore temporality in relation to revision behaviors using keystroke-logging software. Fifty-four L2 English writers composed an argumentative essay and a summary while their keystrokes were logged. Similar to Barkaoui (2019), the researcher divided the writing periods into three periods. Overall, revisions took place during the middle period most frequently. However, when the location of revisions was considered, somewhat different trends emerged. Precontextual revisions (i.e., at the point of inscription) were observed with greater frequency in the middle, whereas contextual revisions

(i.e., away from the point of inscription) took place more often in the last period of writing. In a study of 32 L2 writers of Chinese, Lu and Révész (2021) found similar patterns. Participants completed two narrative and two argumentative writing tasks using the Pinyin typing method and engaged in a stimulated recall after their last performance. The resulting keystroke logs were divided into five periods based on participants' total writing time. Precontextual revisions occurred more frequently in the three middle periods, whereas the incidence of contextual revisions increased from initial to later periods. The stimulated recall comments revealed greater focus on language than content regardless of period, but the number of content-related comments gradually decreased over time. Contrary to the patterns observed by Barkaoui (2016) and Lu and Révész (2021), Gánem-Gutiérrez and Gilmore (2018) reported a steady amount of revision across five periods of writing. The researchers employed eye-tracking, stimulated recall, and screen capture to investigate the composing processes of 22 L2 learners of Japanese. However, this study did not make a distinction between precontextual and contextual revisions, which might have masked some differences in revision behaviors.

Taken together, previous research has yielded ample evidence that L2 writing is a dynamic process, with different cognitive activities dominating various periods during the course of writing. L2 writers appear to focus on planning in initial periods of the writing process, as reflected in verbal protocol comments and fewer but lengthier pauses recorded by keystroke-logging software. According to verbal protocol data, linguistic encoding processes primarily take place during the middle periods. A greater incidence of contextual, local revisions and more frequent, shorter pauses observed during the middle periods are also consistent with a focus on linguistic encoding (Barkaoui, 2016; Lu & Révész, 2021). Finally, some verbal protocol studies found that the main emphasis is on monitoring toward the end of the composing process. This is aligned with the observation that precontextual revisions are more frequent during the final periods of writing (Barkaoui, 2016; Lu & Révész, 2021). These patterns for monitoring, however, were not attested in some studies (e.g., Gánem-Gutiérrez & Gilmore, 2018; Tillema, 2012).

Working memory and the temporal distribution of writing behaviors

Considering the findings of previous research on the time course of writing and Kellogg's (1996, 2001) predictions about the involvement of WM in different writing processes, we would anticipate that various components of WM will play a distinct role at different points during writing. The phonological loop will probably be more implicated in middle and end periods, as it is assumed to be involved in linguistic encoding and monitoring processes. This is expected to be mirrored in stronger links of phonological short-term memory to pausing at lower textual units and all levels of revision toward the middle and end of the writing process. The visual-spatial sketchpad, on the other hand, will likely be called on more at the beginning and end of the writing process, when planning and editing activities are anticipated to dominate to a greater extent. In turn, this will probably be mirrored in stronger relationships of visual-spatial short-term memory to pausing at higher textual units and revisions of larger units in early and late writing periods. Unlike the two slave systems, the influence of the central executive should be present at each period of writing, given its presumed engagement in all subprocesses of writing. By extension, we would expect links between executive functions and all types of pausing and revision during the whole writing process. In

other words, the temporal distribution of pausing and revision is less likely to depend on executive functions.

Although empirical research has not yet tested these predictions, there is a growing amount of research indicating that WM plays a role in L2 writing. Most previous research has been concerned with the relationship of WM to writing outcomes (see Kormos's and Li's contributions to this special issue for more extensive reviews, and Manchón et al.'s contribution for a study of WM effects on CAF measures). More relevant for our current purposes, a small amount of research has also considered the relationship between WM and writing behaviors (e.g., Kim et al., 2021), with two studies investigating these links with respect to several WM memory components and pausing at various locations/revision behaviors at different levels. Vallejos (2020) examined the extent to which WM influenced the length and frequency of pausing by 33 emergent English-Spanish bilinguals. The participants wrote two argumentative essays, one in their L1 English and one in their L2 Spanish, while their keystrokes were logged. The researcher defined two pause thresholds (200 ms and 2 s) and categorized pauses by location (e.g., within words, between sentences). The WM battery included tests of phonological short-term memory (nonword span test), visual-spatial short-term memory (Corsi block task), and executive functions of updating ability (automated operation span) and task-switching (color shape task). The results revealed different patterns for English (L1) and Spanish (L2) writing. In English, the visual-spatial short-term memory scores were found to be related to pause frequency at 200 ms. Visual-spatial short-term memory was also linked to pause frequency within words in Spanish at 200 ms. For Spanish, updating ability was additionally found to have significant correlations with pause length between sentences at 200 ms and pause frequency between sentences at 200 ms. These results point to the importance of taking language proficiency into account when studying the role of WM in writing.

Révész, Michel, et al.'s (2017) previously discussed study also looked into how L2 pausing and revision behaviors might differ depending on writers' WM capacity. Thirty Mandarin L2 users of English carried out an argumentative writing task, during which their keystrokes and eye movements were recorded. Participants were administered a large battery of WM tests, including measures of phonological short-term memory, visual-spatial short-term memory, and various executive functions. The pause threshold was 2 s, and, as in Vallejos' (2020) work, pauses were categorized according to location. Revisions were classified by the level of textual unit changed by the writer (e.g., word, sentence). Three significant correlations were found between WM and writing behaviors: participants with better task-switching ability paused for shorter periods between sentences, those who had superior updating skills paused less frequently between paragraphs, and those with less-developed visual short-term memory viewed the instructions more frequently when they paused.

In summary, both Révész, Michel, et al. (2017) and Vallejos (2020) found that, contrary to expectations, phonological short-term memory did not influence pausing and/or revision behaviors. Yet, as expected, pausing behaviors varied depending on writers' visual short-term memory and executive control. However, the exact nature of the significant patterns observed differed, except for a link between executive control and pausing between sentences. One reason for the nonuniform patterns could be that previous research has not considered the temporal distribution of writing behaviors when examining links between WM and pausing and revision

behaviors. Crucially, as discussed previously, there are theoretical and empirical reasons to believe that, depending on writing period, WM will differentially relate to pausing and revision. The purpose of the present study was to empirically test this hypothesis.

Research Questions

Guided by previous theoretical and empirical work, this study set out to investigate the following research questions:

Research Question 1:

a: To what extent do L2 writers display differential pausing behaviors and pausing-related viewing behaviors at different periods during writing?
b: Do these relationships depend on their phonological short-term memory, visual short-term memory, and/or executive functions?

Research Question 2:

a: To what extent do L2 writers display differential revision behaviors and revision-related viewing behaviors at different periods during writing?
b: Do these relationships depend on their phonological short-term memory, visual short-term memory, and/or executive functions?

In the present study, pausing behaviors were operationalized in terms of length and location (within word, between words, or between sentences) of pauses. Revision behaviors were defined based on the location of revisions (below word, at word level, below clause, clause level or above, or sentence level and above). Viewing behaviors were coded in terms of the syntactic unit (e.g., word, phrase, or sentence) that the participant had viewed while pausing or before revising.

Method

Design

The current data set was collected as part of a broader project examining the links between writing processes, WM capacity, and text quality (see Révész, Michel, et al., 2017; Révész et al., 2019 for reports on other aspects of the project). As part of the present study, we examined 30 L2 writers' performance on Task 2 from the IELTS Academic Writing Test. We recorded their writing behaviors by the means of the keystroke-logging software *Inputlog 6.1.5* (Leijten & Van Waes, 2013) and a *Tobii X2-60* mobile eye tracker. We administered all participants a battery of WM tests.

Participants

The 30 participants were Mandarin first-language speakers and L2 users of English. They were all studying at a university in the United Kingdom, enrolled in masters' ($n = 24$), doctoral ($n = 5$), or bachelor's courses ($n = 1$). Their IELTS overall scores were 7 or higher, corresponding to C1 or higher levels in the Common European Framework of Reference (CEFR). Most of the participants were female ($n = 27$), and the age range was between 18 and 34 ($M = 26.60$, $SD = 3.69$).

Instruments and procedures

Writing task

The writing performances were elicited with a computer-delivered version of Task 2 from the IELTS Academic Writing Test. Participants addressed the following essay prompt:

> Going overseas for university study is an exciting prospect for many people. But while it may offer some advantages, it is probably better to stay home because of the difficulties a student inevitably encounters living and studying in a different culture.

> To what extent do you agree or disagree with this statement? Give reasons for your answer and include any relevant examples from your knowledge or experience.

> Write at least 250 words.

This writing prompt was considered suitable for the participants in terms of topic given that they were all international students studying overseas. Also, we assumed that an argumentative task would pose considerable reasoning demands on writers, making it more likely that any effects of WM emerge (McCormick & Sanz, 2022). Participants were given 40 min to write their essay. They composed in a Microsoft Word document, using size 16 monospace Consolas font type and 1.5 point spaces between lines to enable more accurate eye-gaze measurement.

Working memory tests

We assessed three components of Baddeley's (1986) model of WM: phonological short-term memory, visual short-term memory, and executive control. We evaluated phonological short-term memory (PSTM) with a Mandarin nonword-span (NW) and a Mandarin digit-span test (DS). We assessed visual short-term memory (VSTM) with the forward Corsi block (CF) Task. We measured executive skills with the backward Corsi block (CB), operation-span (OSPAN), color–shape (CS), and stop-signal (SS) tasks. We administered the WM tests in a counterbalanced order across participants.

Both of our PSTM tests, the nonword-span test and digit-span test, were adopted from Zhao's (2013) work. The nonword-span test included 48 one-syllable Chinese nonwords, all of which could be pronounced but had no corresponding characters in Chinese. The nonwords were presented at a rate of one word per second in a random order containing sequences of two to nine nonwords. For each sequence length, the test included three trials. The test began with a brief practice including sequences with two- and three-word nonwords. The longest sequence for which participants could recall at least one sequence correctly was defined as their nonword span. The digit-span test had the same design as the nonword test and was evaluated in the same way. The only difference was that, instead of nonwords, participants were asked to recall two- to nine-digit sequences, randomly generated of numbers from 11 to 99.

The forward Corsi block task, our measure of visual short-term memory, was administered through Inquisit Lab 4. It involved patterns of nine blocks appearing on the computer screen. As part of each trial, two to nine blocks were highlighted, and the participants' task was to click the blocks in the order in which they had been highlighted. The number of blocks highlighted increased from two to nine, with two

trials included for every sequence length. The score for the task was the highest number of blocks that the participants could correctly recall for at least one of the trials.

The backward Corsi block task was employed to measure the executive function of updating ability in a visual context. It had the same design as the forward Corsi block test, with the exception that the participants needed to click the blocks in the reverse sequence as compared with how they had previously seen them highlighted.

The automated operation-span task was used as an additional test of executive control to assess updating ability. First, participants were presented with a math operation on the screen that they had to solve as quickly and accurately as possible, followed by an English letter. This was repeated from three to seven times, after which participants were asked to click the letters in the same order as they had previously appeared. The test contained three sets for each set size, with a set being defined as the number of letters participants had to recall. Different set sizes were presented in a random sequence. For the math operations, an 85% accuracy rate was used as a criterion (Unsworth et al., 2005). As an index, we employed the absolute OSPAN score, which is calculated based on those sets only for which participants could recall all letters accurately.

The color–shape task (Miyake et al., 2004) assessed the executive function of task switching, which we also administered via Inquisit Lab. Participants were asked to assess the shape (e.g., triangle vs. circle) or the color (e.g., red vs. green) of a stimulus including colored shapes. In nonswitching blocks, the participants' task only involved deciding about the shape or the color. In switching blocks, however, they had to evaluate either the shape or the color of the stimulus according to a cue (C or S). Switching cost was defined as the difference in mean reaction times between the two switching and two nonswitching blocks (e.g., Miyake et al., 2004). We trimmed reaction times to exclude values above and below two standard deviations of the mean.

We used the stop-signal task as a measure of inhibitory control, another executive function, which was again presented through Inquisit Lab. An arrow stimulus was shown on the screen, and the participants' task was to press the key "D" in case the arrow pointed to the left and the key "K" if the arrow pointed to the right. However, participants were asked not to respond if the arrow was presented simultaneously with an auditory beep signal. We assessed inhibitory control with mean reaction time (Congdon et al., 2012; Enticott et al., 2006), which captured the amount of time needed for participants to inhibit their response after the auditory signal was presented. We calculated this measure after trimming reaction times to two standard deviations above or below the mean.

Data collection

All participants attended one individual session for which they received a monetary reward in form of a voucher. The completion of the writing test and the WM tests lasted about 2.5 hr. Participants first gave informed consent, and they then completed a brief background questionnaire. Next, we calibrated their eye movements. We used a mobile Tobii X2-60 with a temporal resolution of 60 Hz, mounted to a 23" screen. The participants were seated approximately 60 cms from the center of the screen. We employed a nine-point calibration grid and presented the experiment through Tobii Studio 3.0.9 software (Tobii Technology, n.d.). When the eye calibration was completed, we asked participants to perform the IELTS task and a typing test. After a short break, participants carried out the WM tests. A subset of the participants ($n = 12$) also

engaged in a stimulated recall session prior the WM tests (see Révész, Kourtali, et al., 2017; Révész et al., 2019 for results).

Data analysis

Our data analysis started with dividing the total time each participant spent on the writing task into five equal periods. This allowed us to study how writing behaviors varied across periods between and within participants. Splitting the writing process into five (e.g., Tillema, 2012) instead of three periods (e.g., Roca de Larios et al., 2008) enabled us to obtain a more elaborate picture of how behaviors and associated cognitive processes change during the writing process. To be able to address our research questions, we obtained all indices of writing behaviors for the five periods separately. We calculated the frequency of pauses, revisions, and viewing behaviors using time (per minute) as the denominator.

Analysis of keystroke logs

We employed the keystroke-logging software Inputlog to identify pauses in the data set, adopting a pause threshold of 2 s (e.g., Wengelin, 2006; see, however, Van Waes & Leijten, 2015). We categorized pauses into within-word, between-word, or between-sentence pauses depending on their position. Between-word pauses were considered as one pause, as they frequently included one pause before the press of the spacebar and one pause prior to the start of the subsequent word. We obtained measures of pause length and pause frequency by location.

We also used Inputlog to identify revisions in the keystroke logs. Next, we coded revisions manually according to whether they concerned a below-word-level, a word-level, a below-clause-level, a below-sentence-level, or a sentence-level-and-above change. A second researcher coded 10% of the data, randomly selected, resulting in a high intercoder agreement of 96%.

Analysis of eye-tracking data

We coded the eye-gaze behaviors qualitatively by reviewing participants' eye-gaze behaviors during pauses and prior to revision. First, we identified pauses of 2 s or more and revisions in the keystroke logs, and we then inspected the eye-gaze recordings in Tobii Studio 3.0.9 software to find the same pauses and revisions in the recordings. After the pauses and revisions in the keystroke logs and eye-gaze recordings had been paired, we categorized participants' eye-movements through visual inspection of the eye-gaze data.

For pauses, eye movements were categorized according to whether they stayed for the duration of the pause at the inscription point or visited areas within the word/phrase, clause, sentence, or paragraph appearing immediately before the inscription point. Our coding was dichotomous—that is, we only considered whether a fixation occurred or did not occur in a certain area when participants paused. For each pause, we used as the code the largest textual unit at which participants fixated. For instance, when fixations occurred on a point(s) both outside and within the preceding clause while appearing in the preceding sentence, we coded this series of fixations as "sentence."

For revisions, we inspected eye-gaze behaviors prior to the revision being made, examining whether the writer fixated on an area/areas within the word/phrase, the clause, the sentence, or the paragraph preceding the inscription point. Parallel to

pausing, we employed dichotomous coding, focusing only on whether a fixation was or was not detected within an area prior to revision. We specified the code as the largest textual unit writers viewed before making a revision. For example, when a participant fixated on an area(s) in the preceding word/phrase and beyond while the fixations remained in the preceding clause, this set of fixations was categorized as "clause."

At times, participants viewed the instructions, gazed elsewhere on the screen, or did not gaze at the computer screen when they paused or before they made a revision. We coded these instances, respectively, as "instruction," "elsewhere," and "off-screen." We did not include these categories in our further analyses, given the very small number or absence of observations for these categories for a considerable number of participants at several periods of writing.

Statistical analyses

After calculating descriptive statistics for all our measures of interest, we computed Pearson correlations among participants' performance on the WM tests. To address the research questions, we constructed a series of linear mixed-effects models using the function *lmer* in the R statistical environment. We used the log-transformed values for the pausing and revision indices given the skewed nature of the distributions (the tables for descriptive statistics and figures, however, are based on raw values). We opted for the use of mixed-effects models given the multilevel nature of the data set (we calculated measures for five periods for each participant). The random effect in the models was participant. When addressing research questions 1a and 2a, the fixed effect was writing period alone. To address research questions 1b and 2b, we added a measure of WM and its interaction with writing period as fixed effects. Our predictor of interest was the interaction; a significant interaction would mean that participants behaved differently at various periods of writing depending on their WM capacity. We used the *r.squared GLMM* function in the *MuMIn* package to compute effect sizes for the *lmer* models. Specifically, we obtained marginal and conditional R^2 values (R^2_{m}, R^2_{c}) to assess the variance explained by the fixed effects only and fixed plus random effects together, respectively. We set the alpha level at .01 given the large number of tests we conducted.

Results
Preliminary analyses
Descriptive statistics for and relationships between WM scores
The descriptive statistics for the WM tests are presented in Table 1 (also provided in Révész, Michel, et al., 2017). For each measure, the means and standard deviations indicate that there was enough variance among participants to detect potential WM effects. It is worth noting, however, that the variance for the digit-span and stop-signal tasks was less considerable as compared with the other WM measures, making it somewhat less likely that we would identify significant effects for these measures.

To establish the relationships between the various WM measures, we ran a series of Spearman correlations (see Table 2). The analyses yielded medium-sized correlations (Plonsky & Oswald, 2014) between the Corsi block forward and backward, the nonword-span and color–shape, and the digit-span and stop-signal test results. The rest of the correlations found no significant links. Given that no strong correlations were observed among the various WM measures, we decided to conduct separate mixed-effects analyses for each WM index.

Table 1. Descriptive statistics for working memory measures

	n	M	SD	95% CI lower	95% CI upper
Phonological short-term memory					
Nonword span	29	3.34	1.26	2.90	3.79
Digit span	29	3.83	.71	3.59	4.10
Visual-spatial short-term memory					
Corsi block forward	30	58.80	22.69	50.57	66.76
Executive control					
Corsi block backward	30	57.53	12.42	53.30	61.77
Operation span task	30	51.33	18.37	44.97	58.37
Color–shape task (ms)	30	461.79	297.15	356.51	577.96
Stop-signal task (ms)	30	299.85	58.17	277.10	320.34

Table 2. Spearman correlations among various working memory measures

		Digit span	Corsi forward	Corsi backward	Operation span	Color shape	Stop signal
		r	r	r	r	r	r
	N	p	p	p	p	p	p
Nonword span	29	–.09	.36	–.03	–.21	.44	.02
		.64	.05	.87	.27	.02	.92
Digit span	29		–.11	.05	.23	–.06	–.42
			.57	.81	.23	.76	.02
Corsi forward	30			.47	.15	–.20	.02
				.01	.44	.29	.92
Corsi backward	30				.03	–.34	.20
					.86	.07	.29
Operation span	30					–.30	.05
						.10	.80
Color shape	30						.06
							.76

Descriptive statistics for writing behaviors

The descriptive statistics for pausing behaviors and pausing-related viewing behaviors are provided in Tables 3 and 4, respectively. Similar to the WM indices, the means and standard deviations show that there was enough variance among participants to address the effects of writing period and WM on pausing behaviors and eye-gaze behaviors during pauses.

The descriptive statistics for revision behaviors and revision-related viewing behaviors are given in Tables 5 and 6, respectively. As for the WM and pausing-related measures, the means and standard deviations indicate that the variance among participants was sufficiently large to address the influence of writing period and WM on revision behaviors and eye-gaze behaviors prior to revision.

Research question 1: Writing period, working memory, and pausing behaviors

Research question 1a investigated the effects of writing period on pausing behaviors and eye-gaze behaviors during pauses. To address this question, we conducted a series of linear mixed-effects analyses. In each analysis, our fixed effect was writing period, we

Table 3. Descriptive statistics for pausing behaviors (*N* = 30)

	M	SD	95% CI lower	95% CI upper
Pause frequency per minute				
Period 1				
Within words	0.26	0.27	0.17	0.37
Between words	2.03	0.66	1.81	2.26
Between sentences	0.40	0.31	0.30	0.51
Period 2				
Within words	0.30	0.27	0.21	0.40
Between words	2.13	0.65	1.89	2.37
Between sentences	0.38	0.31	0.27	0.49
Period 3				
Within words	0.34	0.23	0.27	0.42
Between words	2.42	1.09	2.07	2.80
Between sentences	0.29	0.33	0.18	0.41
Period 4				
Within words	0.21	0.19	0.14	0.28
Between words	2.23	0.83	1.94	2.50
Between sentences	0.22	0.23	0.15	0.31
Period 5				
Within words	0.21	0.21	0.15	0.29
Between words	1.65	0.99	1.31	2.01
Between sentences	0.15	0.16	0.09	0.20
Median Pause Length				
Period 1				
Within words	3.83	5.30	2.31	5.92
Between words	3.87	1.40	3.39	4.36
Between sentences	14.31	15.89	8.90	19.93
Period 2				
Within words	3.12	3.25	2.12	4.28
Between words	4.05	1.25	3.64	4.57
Between sentences	4.03	3.45	2.83	5.26
Period 3				
Within words	2.94	2.32	2.19	3.79
Between words	3.63	0.76	3.36	3.87
Between sentences	3.27	3.38	2.17	4.60
Period 4				
Within words	2.70	2.27	1.93	3.54
Between words	3.74	1.22	3.35	4.20
Between sentences	4.01	5.14	2.31	5.91
Period 5				
Within words	2.49	2.31	1.72	3.33
Between words	3.05	1.21	2.60	3.45
Between sentences	1.97	2.57	1.12	2.93

included a random intercept for participant, and the dependent variable was one of the pausing-related measures. Out of 50 analyses we carried out, writing period emerged as a significant predictor for three indices: pause frequency between words, median pause length between words, and median pause length between sentences. Table 7 summarizes and Figure 1 illustrates the significant relationships we found (see also Table S1 in Supplementary Information Online for the full model results).

Participants made significantly fewer and shorter pauses between words during Period 5 as compared with all previous periods. Pauses were also significantly longer between sentences during Period 1 compared with all later periods of writing and during Period 2 compared with Period 5. The effect size was the largest for pause length

Table 4. Descriptive statistics for presence of eye-gaze during pause per minute (*N* = 30)

	M	*SD*	95% CI lower	95% CI upper
Period 1				
Word/phrase	0.72	0.46	0.58	0.90
Clause	0.88	0.63	0.67	1.13
Sentence	0.56	0.52	0.39	0.76
Paragraph	0.65	0.41	0.51	0.79
Period 2				
Word/phrase	0.59	0.50	0.41	0.76
Clause	0.78	0.68	0.55	1.02
Sentence	0.49	0.40	0.35	0.64
Paragraph	1.00	0.60	0.79	1.20
Period 3				
Word/phrase	0.58	0.54	0.40	0.78
Clause	0.91	0.83	0.64	1.22
Sentence	0.41	0.28	0.31	0.51
Paragraph	0.85	0.59	0.67	1.06
Period 4				
Word/phrase	0.66	0.56	0.47	0.86
Clause	0.75	0.62	0.54	0.99
Sentence	0.49	0.38	0.36	0.63
Paragraph	1.01	0.61	0.80	1.23
Period 5				
Word/phrase	0.63	0.65	0.43	0.86
Clause	0.74	0.63	0.52	0.99
Sentence	0.44	0.34	0.33	0.57
Paragraph	1.01	0.91	0.73	1.34

between sentences, followed by pause frequency between words and median pause length between words in this order, explaining 21%, 12%, and 8% of the variation, respectively.

Research question 1b examined whether WM influenced the extent to which writing period related to participants' pausing behaviors and eye-gaze behaviors during pauses. To investigate this question, we conducted another series of linear mixed-effects analyses. Participant served as the random effect; the fixed effects included a WM measure, writing period, and their interaction; and the dependent variable was a pausing or pause-related eye-gaze behavior index. A significant interaction would mean that, depending on participants' WM, writing period had a differential relationship with the pausing-related measure in the model. In other words, when we observed a significant interaction effect, the relationship between writing period and pausing varied across participants with different WM. We ran each model for all five periods as reference points to identify all the possible interactions between WM and period of writing. The analyses yielded a significant interaction effect for five pausing indices, including six types of WM measures. Table 8 provides the significant interactions identified, and Figures 2 and 3 illustrate these relationships (the full results for the models are available in Tables S2–S6 in the Supplementary Information Online).

For within-word pausing, a significant interaction was observed with the CS scores. Participants with better CS reaction times paused more frequently during within-word pauses at Period 3 than those with lower CS reaction times, whereas performance on the color shape test did not appear to have much influence on within-word pause frequency during Period 1.

Table 5. Descriptive statistics for frequency of revision per minute ($N = 30$)

	M	SD	95% CI lower	95% CI upper
Period 1				
Below word	2.76	2.63	2.06	3.88
Single word	0.96	1.11	0.66	1.43
Below clause	0.85	0.91	0.55	1.20
Clause & above	0.03	0.09	0.01	0.07
Sentence & above	0.03	0.07	<0.01	0.05
Period 2				
Below word	2.96	2.53	2.25	3.97
Single word	1.53	2.20	0.97	2.49
Below clause	1.19	1.39	0.74	1.72
Clause & above	0.08	0.17	0.03	0.14
Sentence & above	0.09	0.19	0.02	0.16
Period 3				
Below word	3.41	3.61	2.34	4.84
Single word	1.43	1.79	0.92	2.16
Below clause	1.31	1.42	0.84	1.88
Clause & above	0.09	0.14	0.04	0.14
Sentence & above	0.05	0.12	0.02	0.10
Period 4				
Below word	2.68	2.95	1.77	3.70
Single word	1.32	1.76	0.82	2.01
Below clause	1.63	1.71	1.09	2.24
Clause & above	0.13	0.20	0.07	0.20
Sentence & above	0.07	0.16	0.02	0.13
Period 5				
Below word	3.27	3.90	2.19	4.97
Single word	1.76	2.53	1.11	2.78
Below clause	1.57	1.35	1.13	2.10
Clause & above	0.15	0.25	0.07	0.24
Sentence & above	0.16	0.28	0.07	0.27

Turning to between-word pause frequency, several significant interactions were identified between pausing and WM involving nonword-span and Corsi block forward scores. Although participants' nonword spans did not have a notable relationship with pause frequency during Periods 1 and 2, at Period 5 participants with higher nonword spans paused more often between words. At Period 4, participants showed a pattern similar to that in Period 5, making participants' pausing behavior at Periods 4 and 5 significantly different from that at Period 3, where lower non-word-span participants paused more often between words. At Period 3, participants with lower forward Corsi block scores also showed greater between-word pause frequency as compared with Period 5 where the opposite pattern was observed.

Pause length within words was found to vary between Periods 2 and 4 depending on participants nonword span scores. Participants with higher nonword spans paused longer within words at Period 2 but displayed shorter pauses at Period 4.

For pause length between words, we found significant interactions between period of writing and participants' Corsi forward and backward scores. During Period 1, participants with higher Corsi forward scores paused shorter between words. This trend was significantly different from Periods 3 and 4, where the Corsi forward scores appeared to have little influence on pause length. The Period 1 pattern also differed significantly from Period 2, where participants with higher Corsi forward spans displayed longer pauses between words. For Corsi backward, we found a significant interaction between

Table 6. Descriptive statistics for presence of eye-gaze before revision per minute ($N = 30$)

	M	*SD*	95% CI lower	95% CI upper
Period 1				
Word/phrase	1.62	1.16	1.25	2.06
Clause	0.07	0.11	0.03	0.11
Sentence	0.20	0.18	0.14	0.27
Paragraph	0.06	0.13	0.02	0.11
Period 2				
Word/phrase	1.68	1.40	1.23	2.19
Clause	0.15	0.21	0.08	0.24
Sentence	0.23	0.29	0.14	0.34
Paragraph	0.17	0.25	0.09	0.25
Period 3				
Word/phrase	1.60	1.47	1.13	2.16
Clause	0.15	0.26	0.07	0.25
Sentence	0.31	0.37	0.18	0.44
Paragraph	0.18	0.35	0.07	0.32
Period 4				
Word/phrase	1.42	1.34	0.97	1.91
Clause	0.22	0.46	0.10	0.39
Sentence	0.32	0.34	0.21	0.44
Paragraph	0.11	0.29	0.03	0.22
Period 5				
Word/phrase	1.37	1.24	0.94	1.83
Clause	0.18	0.34	0.07	0.30
Sentence	0.40	0.87	0.16	0.75
Paragraph	0.37	0.76	0.13	0.67

Table 7. Significant time effects identified for pausing behaviors

Dependent variable/Ref level	Predictor	*Est*	*SE*	*t*	*p*	R^2_m	R^2_c
Pause frequency between words per minute							
Period 1	Period5	−0.17	0.05	−3.76	<.01		
Period 2	Period5	−0.19	0.05	−4.26	<.01		
Period 3	Period5	−0.22	0.05	−4.99	<.01		
Period 4	Period5	−0.19	0.05	−4.25	<.01	.12	.41
Median pause length between words							
Period 1	Period5	−0.10	0.03	−3.20	<.01		
Period 2	Period5	−0.12	0.03	−4.04	<.01		
Period 3	Period5	−0.08	0.03	−2.72	.01		
Period 4	Period5	−0.08	0.03	−2.86	.01	.08	.35
Median pause length between sentences							
Period 1	Period2	−0.37	0.09	−3.96	<.01		
	Period3	−0.46	0.09	−4.88	<.01		
	Period4	−0.44	0.09	−4.72	<.01		
	Period5	−0.61	0.09	−6.49	<.01		
Period 2	Period5	−0.24	0.09	−2.53	.01	.21	.32

Periods 1 and 5, participants with higher Corsi backward scores showing shorter pauses in Period 1 but longer pauses in Period 5.

Moving onto median pause length between sentences, the mixed-effects analyses yielded significant interactions between pausing and the nonword-span and OSPAN results. Participants with greater nonword scores paused considerably longer between sentences in Period 1, whereas they paused somewhat shorter in Period 2. Similar
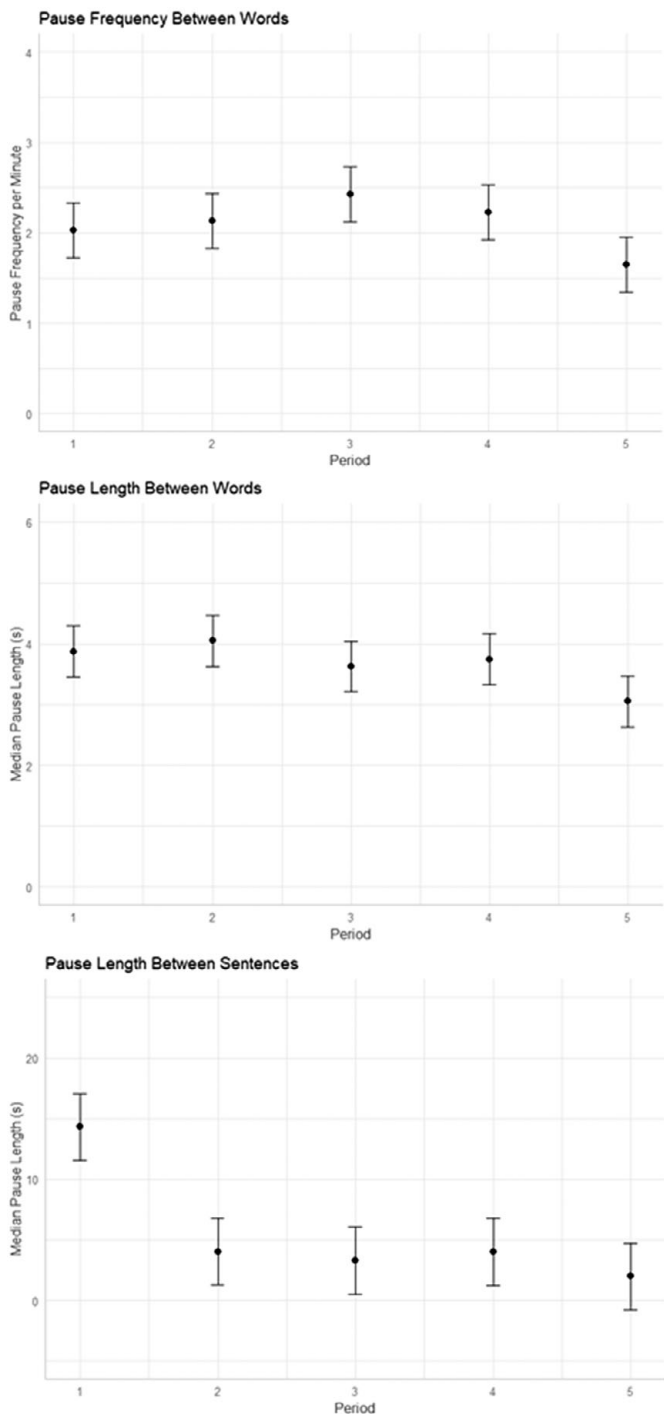
Figure 1. Significant period effects: Pausing behaviors.

Table 8. Significant working memory by period interaction effects identified for pausing behaviors

Dependent variable/Ref level	Predictor	Est	SE	t	p
Pause frequency within words					
Period 1	CS:Period3	<0.01	<0.01	−2.54	.01
Pause frequency between words					
Period 1	NWS:Period5	0.10	0.03	3.06	<.01
Period 2	NWS:Period5	0.11	0.03	3.26	<.01
Period 3	NWS:Period4	0.11	0.03	3.33	<.01
	NWS:Period5	0.15	0.03	4.57	<.01
	CF:Period5	0.01	<0.01	2.66	.01
Median pause length within words					
Period 2	NWS:Period4	−0.15	0.06	−2.60	.01
Median pause length between words					
Period 1	CF:Period2	<0.01	<0.01	3.60	<.01
	CF:Period3	<0.01	<0.01	2.67	.01
	CF:Period4	<0.01	<0.01	3.37	<.01
	CB:Period5	0.01	<0.01	2.51	.01
Median pause length between sentences					
Period 1	NWS:Period2	−0.19	0.07	−2.61	.01
Period 3	OSPAN:Period4	−0.02	<0.01	−3.15	<.01

Note. CS = Color Shape; NWS = Non-word span; CF = Corsi Block forward; CB = Corsi Block backward; OSPAN = Operation span.

patterns were observed for the OSPAN results and Periods 3 and 4, with higher OSPAN participants pausing longer in Period 3 but shorter in Period 4.

Research question 2: Working memory, writing period, and revision behaviors

To address research question 2a, we carried out a series of linear mixed-effects analyses for revision behaviors and eye-gaze behaviors prior to revision. In each model, the fixed effect was period of writing and participant was added as a random intercept. The dependent variable was one of the revision-related indices. Out of the 45 models constructed, four yielded a significant effect for writing period: revision at word level, revision below clause level, revision at clause level, and presence of eye gaze at previous sentence before revision. The significant relationships are summarized in Table 9 and illustrated in Figure 4 (see Table S7 in Supplementary Information Online for the full model results). Overall, participants made more revisions in later than earlier periods of writing. More specifically, we observed more word-level revisions in Periods 3 and 5 than at Period 1, more below-clause revisions in Periods 4 and 5 than Periods 1 and 2, and more clause-level revisions in Period 4 than in Periods 1 and 2. Participants also viewed the previous sentence more frequently during Period 5 than Period 1.

In general, the effect sizes were smaller than for pausing-related behaviors, explaining 4%–9% of the variation (word-level revision: 4%, below-clause revision: 9%, clause-level revision: 6%, view previous sentence before revision: 5%).

To address research question 2b, we ran the same analyses as for research question 1b, the only difference being that revisions behaviors served as the dependent variables in the models. None of the analyses yielded a significant interaction effect, which means that writing period did not have a significant influence on the relationship between participants' revision behaviors and WM.
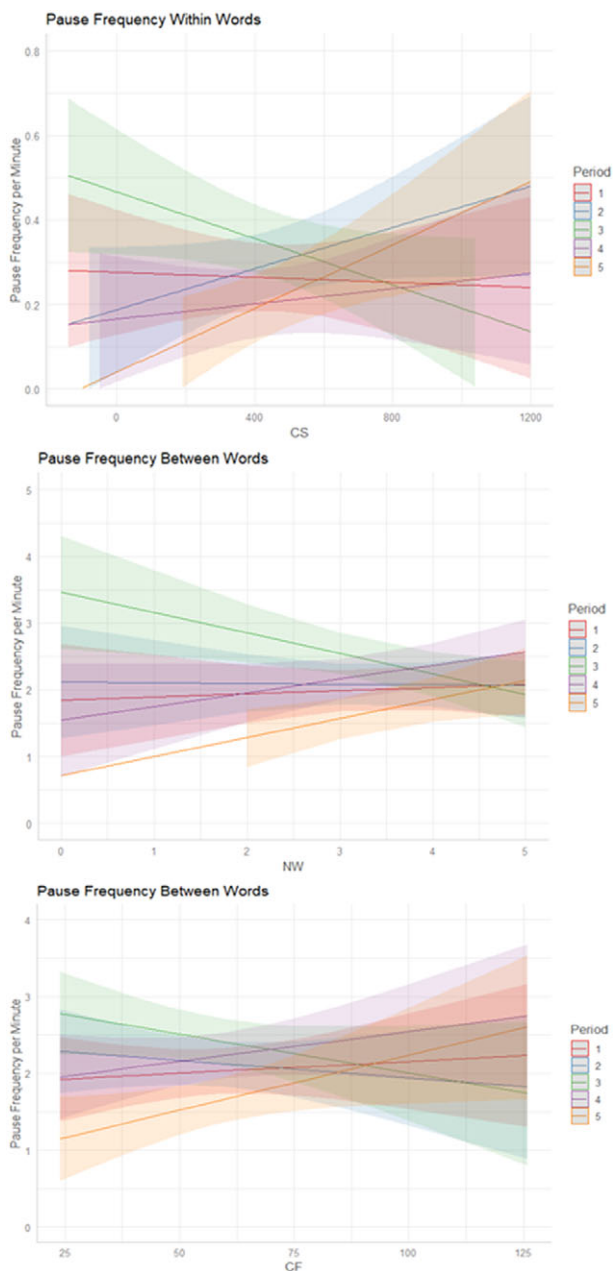
Figure 2. Significant working memory by time interaction effects: Pause frequency.

Summary of results

The results of the study for research questions 1a and 2a are summarized in Table 10 and for research questions 1b and 2b in Table 11.
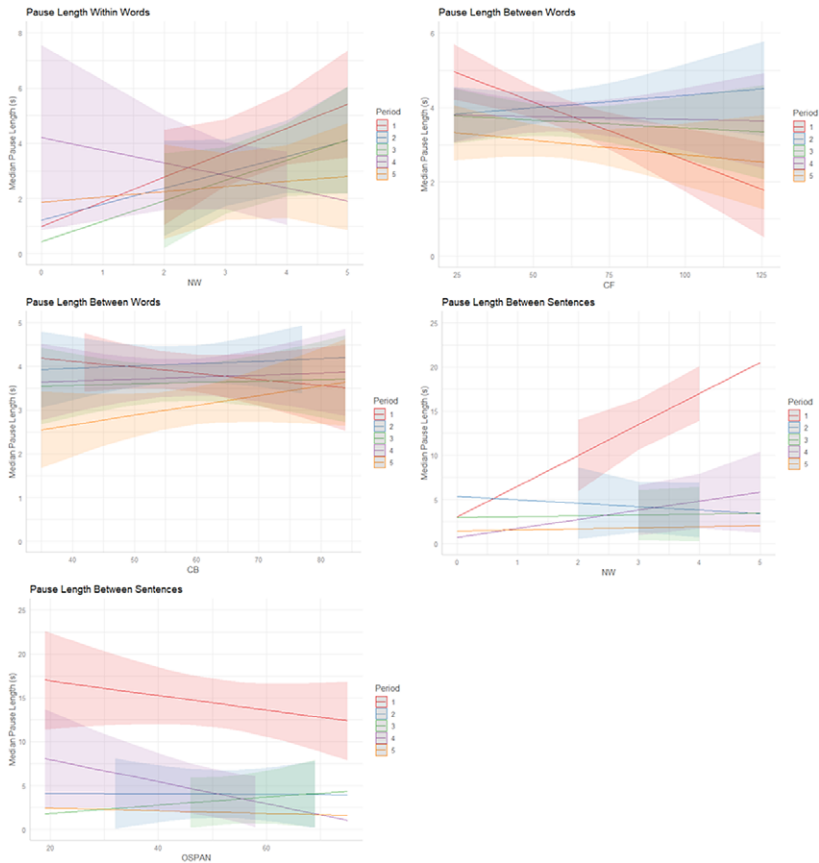
Figure 3. Significant working memory by time interaction effects: Pause length.

Discussion

The first part of our first research question asked the extent to which L2 writers display differential pausing behaviors and pausing-related viewing behaviors at various periods of writing. We found longer between-sentence pauses in the initial periods than in later periods of writing. Our results also revealed that participants made fewer and shorter between-word pauses in the final period of writing than in previous periods. These patterns are aligned with the results of Xu and Qi (2017) and Barkaoui (2019), who also revealed greater length of pausing in the beginning periods of composing by advanced L2 users. The specific trends observed by pause location are also consistent with the broader findings of previous research on writing processes. As between-sentence pauses are likely to be associated with planning processes (e.g., Révész, Kourtali, et al., 2017; Révész et al., 2019; Schilperoord, 1996), the greater incidence of between-sentence pauses in earlier periods of writing suggests more engagement in planning in the early periods of composing. On the other hand, the decreased frequency of between-word pauses in the final period implies a decreased focus on linguistic encoding, as between-word pauses often reflect translation processes (e.g., Révész,

Table 9. Significant period effects identified for revision behaviors and revision-related eye-gaze behaviors

Dependent variable/Ref level	Predictor	Est	SE	t	p	R^2_m	R^2_c
Revision at word level per minute							
Period 1	Period3	0.17	0.06	2.87	<.01		
	Period5	0.18	0.06	3.10	<.01	.04	.56
Revision below clause level per minute							
Period 1	Period4	0.30	0.07	3.99	<.01		
	Period5	0.21	0.07	2.76	.01		
Period 2	Period4	0.31	0.07	4.13	<.01		
	Period5	0.22	0.07	2.90	<.01	.09	.49
Revision at clause level per minute							
Period 1	Period4	−0.25	0.08	−3.12	<.01		
Period 2	Period4	−0.22	0.08	−2.69	.01	.06	.21
Eye-gaze at previous sentence before revision per minute							
Period 1	Period5	0.22	0.08	2.84	.01	.05	.20

Kourtali, et al., 2017; Révész et al., 2019; Schilperoord, 1996). Parallel to these findings, existing studies have, overall, found that planning activities dominate earlier periods of writing and linguistic encoding processes occur more often in the middle of the writing process (Barkaoui, 2015; Khuder & Harwood, 2015; Michel et al., 2020; Roca de Larios et al., 2008; Tillema, 2012; Van Weijen, 2009).

The first research question additionally investigated the extent to which the temporal distribution of pausing behaviors varied according to L2 writers' phonological short-term memory, visual short-term memory, and executive functions. Drawing on Kellogg's (1996, 2001) writing model, prior empirical work on writing periods (which was largely replicated in the present research) and previous work showing a link between pause locations and linguistic encoding processes (e.g., Révész, Kourtali, et al., 2017; Révész et al., 2019; Schilperoord, 1996), we anticipated that pausing behaviors at lower textual units would be more influenced by phonological short term-memory in the middle and toward the end of writing, when linguistic encoding and monitoring processes probably take place more frequently. In line with our expectations, participants' nonword-span scores showed several significant relationships with the frequency and length of participants' pauses within and between words in the middle and end periods.

The specific relationships observed for PSTM, however, paint a complex picture, including some more and less anticipated patterns. Those with lower nonword-span scores produced more between-word pauses toward the middle of the writing process, suggesting that low-PSTM writers, as expected, struggled more with linguistic encoding processes (e.g., Révész, Kourtali et al., 2017; Révész et al., 2019; Schilperoord, 1996). Surprisingly, however, writers with higher nonword-span scores produced longer pauses within words in the middle periods. This might have been due to an enhanced concern with spelling at this period. The opposite trends were observed for later periods of writing, with higher non-word-span scores being associated with increased between-word pausing and lower non-word-span scores linked to longer within-word pausing. The higher number of between-word pausing by higher PSTM participants might be an indicator of greater focus on monitoring at the word level, whereas longer within-word pausing by lower PSTM writers might have captured more attention to below-word-level issues, such as spelling. If so, these patterns possibly resulted from the fact that high-PSTM participants, unlike their low-PSTM counterparts, had already resolved
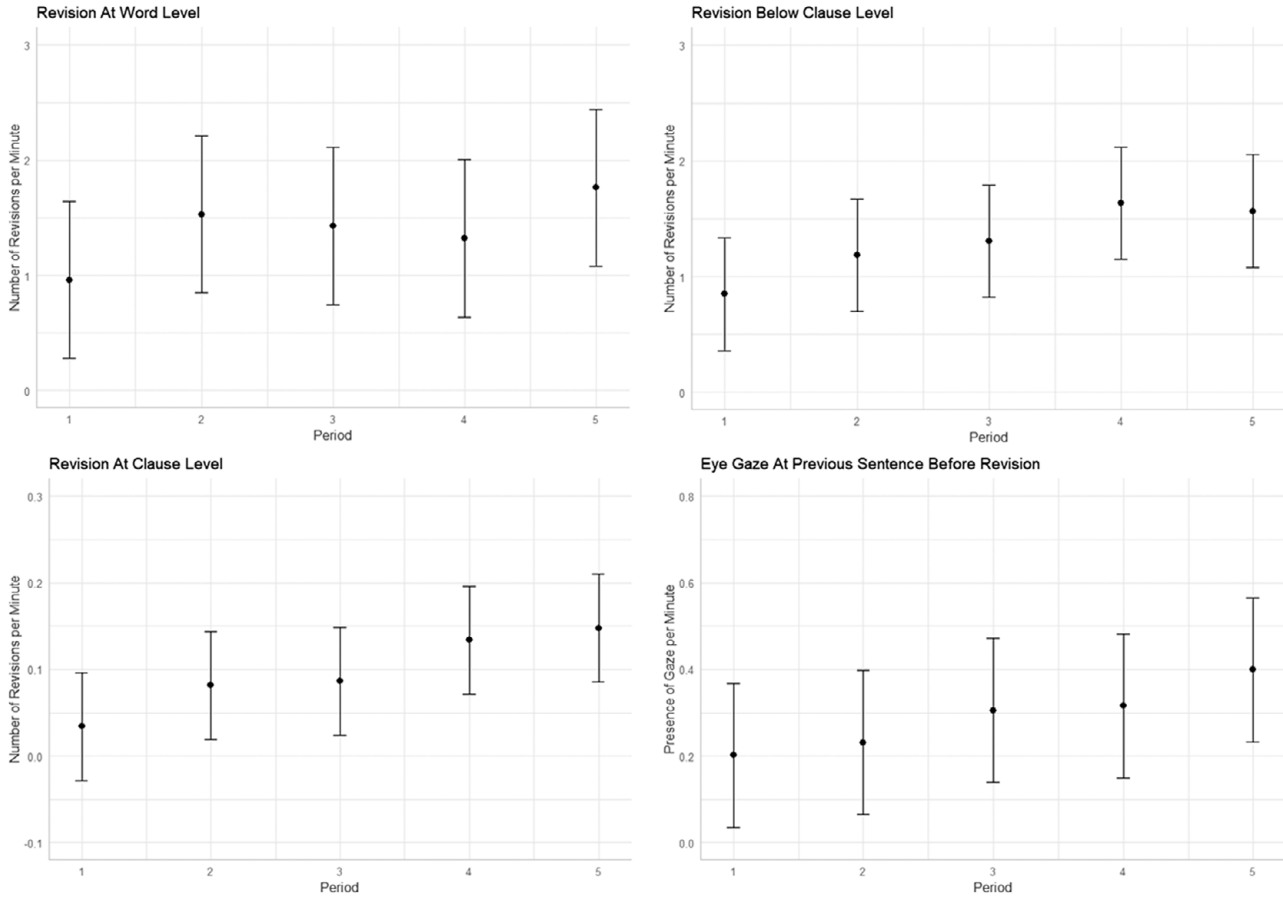
Andrea Révész, Marije Michel and Minjin Lee

Figure 4. Significant period effects: Revision behaviors.

Table 10. Summary of differences across periods

Measure	Significant differences across time periods
Pause frequency between words	P1, P2, P3, P4 > P5
Median pause length between words	P1, P2, P3, P4 > P5
Median pause length between sentences	P1 > P2, P3, P4, P5; P2 > P5
Revision frequency at word level	P1 < P3, P5
Revision frequency below clause level	P1, P2 < P4, P5
Revision frequency at clause level	P1, P2 < P4
Eye-gaze frequency at previous sentence before revision	P1 < P5

Note. P = Period.

Table 11. Summary of WM effects across periods

Measure	WM test	Periods affected
Pause frequency between words	Non-word span ↑	P3 ↓ - P4, P5 ↑
Pause frequency between words		P5 ↑ - P1, P2 ~
Median pause length within words		P2 ↑ - P4 ↓
Median pause length between sentences		P1 ↑ - P2 ↓
Pause frequency between words	Corsi forward ↑	P3 ↓ - P5 ↑
Median pause length between words		P1 ↓ - P2 ↑, P3, P4 ~
Median pause length between words	Corsi backward ↑	P1 ↓ - P5 ↑
Pause frequency within words	Color shape ↑	P3 ↑ - P1 ~
Median pause length between sentences	OSPAN ↑	P3 ↑ - P4 ↓

Note. ↑ means higher/longer, ↓ means lower, ~ means small/no effect.

below-word-level issues (e.g., spelling) during the middle periods of writing, allowing them to allocate more attentional capacity to other linguistic encoding processes (e.g., lexical retrieval) in the final period. Different from what we envisaged, we also found that higher PSTM was associated with longer pauses between sentences, a likely reflection of planning processes, at the beginning of composing. Although previous research has shown that writers do vary as to their planning behaviors (e.g., Cumming, 1989), a tentative explanation for this finding could be that at least some participants with better PSTM skills engaged in deeper content planning with relatively little concern for linguistic issues in the initial period of writing, as they anticipated fewer linguistic difficulties in translating their ideas into linguistic form in later periods of their composing process.

Our predictions for the influence of visual-spatial short-memory on pausing across writing periods were partially confirmed. Based on Kellogg's (1996, 2001) model and prior empirical work, we assumed that visual-spatial short-term memory would play a more prominent role in the initial and final periods of writing. Our rationale for this prediction was that planning and editing activities, processes that are expected to rely on the use of images and visual-spatial information, respectively, are likely to take place with greater frequency in these periods. Indeed, we found more significant links between visual-spatial short-term memory and pausing behaviors for the beginning and end periods than for the middle periods of writing.

Turning to more specific trends, we anticipated stronger links of visual-spatial short-term memory to pausing behaviors at higher textual units. The negative association found between Corsi block scores and pause length in the initial period is unsurprising, as those with better visual-spatial short-term memory were probably more successful in

generating prelinguistic ideas entailing images during planning (Kellogg, 1996, 2001). It was contrary to our prediction, however, that pauses would be shorter between words rather than sentences, as between-word and between-sentence pauses have been posited to be more associated with linguistic encoding and planning processes, respectively (e.g., Révész, Kourtali, et al., 2017; Révész et al., 2019; Schilperoord, 1996). Possibly, writers with better visual-spatial short-term memory were more able to retrieve images associated with concepts in their lexicon, which might have been reflected in the negative association of between-word pauses with visual-spatial short-term memory in the initial period. In the middle periods, participants with lower Corsi block scores were found to pause more often between words; this was probably due to experiencing more difficulty with retrieving the formal properties of words (e.g., spelling) or meaning representations involving images. The final period engaged those with higher visual-spatial short-term memory in longer and more between-word pauses. This pattern is difficult to account for, and further research is warranted to shed light on mechanisms underlying this finding.

We did not anticipate any differential influence of central executive skills depending on writing period, as executive functioning is predicted to be involved throughout the whole writing process to a large degree. Indeed, we found fewer significant relationships for executive functions than for other components of WM. Nevertheless, two links to pausing behaviors emerged. During the middle periods of writing, better scores on the color shape task were associated with a higher incidence of within-word pauses; probably participants with increased task-switching ability more often moved their attention between lower-level and higher-level writing subprocesses, which was captured in increased within-word pausing associated with lower-level linguistic encoding processes. Also, participants who had higher operation span scores produced shorter pauses between sentences toward the end of composing, maybe owing to their better ability to coordinate and update monitoring operations. It is worth noting that previous studies found a similar link between executive functioning and pause length between sentences (Vallejos, 2020) and pause length in general (Kim et al., 2021).

The first part of second research question asked the extent to which L2 writers show differential revision behaviors and revision-related viewing behaviors at various writing periods. Our results revealed a greater amount of revision in the final than during the initial periods of writing. This pattern was uniform for various levels of revision, with the number of word-level, below-word-level, and below-clause-level revisions all increasing over time. Additionally, participants were found to view the previous sentence they had produced less frequently in the beginning than at the end of writing. These findings correspond to the results of some of the previous research (Barkaoui, 2015; Roca de Larios et al., 2008), where revision, as reflected in verbal protocol comments, took place more frequently toward the end of the writing process. The revision trends detected here are also aligned with the results for contextual revisions in Barkaoui's (2016) and Lu and Révész's (2021) study, which investigated revision behaviors by the means of keystroke-logging software. A possible explanation for greater alignment with studies observing growing amount of revision as time progressed might have to do with the relatively high proficiency level of our participants. Roca de Larios et al. (2008), for example, observed greater variation and more strategic distribution of activities by higher than lower proficiency writers over time. Further research on the moderating effect of proficiency in revision behavior is needed to shed more light on this link.

The second research question also addressed whether any relationships between revision behaviors and period of writing would differ depending on L2 writers' phonological short-term memory, visual short-term memory, and executive functions. Based on Kellogg's (1996, 2001) model of writing and previous research on the temporal dimension of writing processes, we predicted that phonological and visual-spatial short-term memory would play a bigger role toward the end of the writing process when monitoring processes are likely to dominate. Like for pausing, we did not anticipate a moderating role for executive skills as a function of writing period, as the central executive, according to Kellogg, is implicated at each writing period. Given that there was considerable variation among participants' phonological and visual-spatial short-term memory spans, it was against our prediction that WM did not influence the time distribution of revision behaviors. A reason for the lack of significant effects may be that the participants in the present study were high-proficiency writers with much academic writing experience. Also, the argumentative writing task, contrary to our expectation, might not have posed heavy cognitive demands on our participants, probably because they were familiar with the IELTS writing task type that they carried out. The combination of high proficiency and manageable task demands might have enabled them to successfully employ writing strategies they had developed throughout their past studies, which, in turn, might have led to decreased cognitive load on their part compensating for potentially lower phonological and/or visual-spatial short-term storage capacity (McCormick & Sanz, 2022). At higher levels of proficiency, revision, a largely conscious process, is probably more susceptible to strategic behavior than other writing behaviors such as pausing at lower textual units. Indeed, the present study yielded the most significant WM links for pausing within and between words as a function of writing period. For high-proficiency writers, pausing at lower textual units is probably underlain by implicit processes to a greater degree than revision, given that pausing within and between words is often associated with linguistic encoding processes that tend to be more automatic as proficiency increases.

Limitations and further research directions

Before drawing our conclusions, it is necessary to consider the limitations of this research. One weakness of this study is that we adopted a single, relatively long pause threshold of 2 s. Although this pause threshold has traditionally been employed in L2 writing research, a shorter threshold would have made it possible to gain a more complete understanding of the influence of writing period and WM on lower-level linguistic encoding processes (Baaijen et al., 2012; Michel et al., 2020; Vallejos, 2020; Van Waes & Leijten, 2015). A second limitation has to do with the relatively low precision of the eye tracker we used. Higher-precision equipment would have allowed us to gain more specific measurements, enabling a more thorough investigation of any effects of writing period and WM on eye-gaze behaviors during the writing process. A third limitation is that we have no reliability estimates available for the WM tests. The study would also have benefited from an even more detailed investigation of the temporal distribution of writing activities. In future research, it would be worthwhile to divide the writing process into even shorter periods to examine more thoroughly the potentially differential effects of WM on writing behaviors over time. Alternatively, researchers could identify writing periods for participants on an individual basis, considering the actual writing behaviors they

display rather than passage of time. This would help make more accurate inferences about links between the temporal distribution of writing behaviors and various components of WM. Further valuable research avenues would include examining whether the results observed here would transfer to other writing task types. As we mentioned earlier, participants were familiar with the argumentative task type investigated in the current study, which might have enabled them to engage in more strategic behavior than other task types would have allowed for. Another worthwhile future research direction would be to investigate whether our results are replicated for different levels of proficiency. Previous research suggests that writing behaviors vary as a function of L2 proficiency (e.g., Révész et al., 2022), and some studies focusing on other areas of L2 competence found that the facilitative effect of cognitive abilities declines with increasing proficiency (e.g., Serafini & Sanz, 2016). It would also be interesting to investigate the potential interaction of other cognitive individual differences such as aptitude with the time distribution of pausing and revision behaviors. The results of this line of research have the potential to provide guidance for developing assistive technologies and outlining guidelines for accommodating L2 writers with varied cognitive skills (Granena, 2023, in this issue; Kormos, 2021, 2023, in this issue; Michel et al., 2019).

Conclusion

In this study, our goal was to investigate how pausing and revision behaviors may differ across the time course of writing and how individual differences in WM may moderate the temporal distribution of pausing and revision behaviors. The pausing and revision patterns we observed are largely consistent with the conclusions of previous research that planning, linguistic encoding, and monitoring processes take place with greater frequency in the initial, middle, and later periods of writing, respectively. Although we did not detect differences in revision behaviors across writing periods depending on writers' WM, we found that various components of WM had differential effects on pausing behaviors during the course of writing. Our results for pausing largely reflected the predictions we derived from Kellogg's (1996, 2001) model of writing and WM, with phonological and visual-spatial short-term memory observed to influence the time distribution of pausing to a considerably greater degree than executive functions.

References

Alves, R. A., Castro, S. L., de Sousa, L., & Stromqvist, S. (2007). Influence of keyboarding skill on pause-execution cycles in written composition. In Torrance, M., Van Waes, L., & Galbraith, D. (Eds.), *Writing and cognition: Research and applications* (pp. 55–65). Elsevier.

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

Baddeley, A. (1986). *Working memory*. Clarendon Press/Oxford University Press.

Baddeley, A. D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, 4, 417–423. https://doi.org/10.1016/S1364-6613(00)01538-2

Barkaoui, K. (2015). Test takers' writing activities during the TOEFL iBT® writing tasks: A stimulated recall study. *ETS Research Report Series*, 2015, 1–42. http://doi.org/10.1002/ets2.12050

Barkaoui, K. (2016). What and when second-language learners revise when responding to timed writing tasks on the computer: The roles of task type, second language proficiency, and keyboarding skills. *The Modern Language Journal*, 100, 320–240. https://doi.org/10.1111/modl.12316

Barkaoui, K. (2019) What can L2 writers' pausing behaviour tell us about their L2 writing processes? *Studies in Second Language Acquisition*, 41, 529–554. https://doi.org/10.1017/S027226311900010X

Chukharev-Hudilainen, E., Feng, H.-H., Saricaoglu, A., & Torrance, M. (2019). Combined deployable keystroke logging and eyetracking for investigating cognitive processes that underlie L2 writing. *Studies in Second Language Acquisition*, 41, 583–604. https://doi.org/10.1017/S027226311900007X

Congdon, E., Mumford, J. A., Cohen, J. R., Galvan, A., Canli, T., & Poldrack, R. A. (2012). Measurement and reliability of response inhibition. *Frontiers in Psychology*, 3, 1–10. https://doi.org/10.3389/fpsyg.2012.00037

Cumming, A. (1989). Writing expertise and second language proficiency. *Language Learning*, 39, 81–141. https://doi.org/10.1111/j.1467-1770.1989.tb00592.x

DeKeyser, R.M. (2012). Interactions between individual differences, treatments, and structures in SLA. *Language Learning*, 62, 189–200. https://doi.org/10.1111/j.1467-9922.2012.00712.x

Enticott, P.G., Ogloff, J. R., & Bradshaw, J. L. (2006). Associations between laboratory measures of executive inhibitory control and self-reported impulsivity. *Personality and Individual Differences*, 41, 285–294. https://doi.org/10.1016/j.paid.2006.01.011

Gánem-Gutiérrez, G. A., & Gilmore, A. (2018). Tracking the real-time evolution of a writing event: Second language writers at different proficiency levels. *Language Learning*, 68, 468–506. https://doi.org/10.1111/lang.12280

Granena, G. (2023). Cognitive individual differences in the process and product of L2 writing. *Studies in Second Language Acquisition*, 45, 765–785.

Hayes, J. R. (1996). A new framework for understanding cognition and affect in writing. In C. M. Levy & S. Ransdell (Eds.), *The science of writing* (pp. 1–27). Erlbaum.

Kellogg, R. (1996). A model of working memory in writing. In M. Levy & S. Ransdell (Eds.), *The science of writing: Theories, methods, individual differences, and applications* (pp. 57–72). Lawrence Erlbaum.

Kellogg, R. (2001). Competition for working memory among writing processes. *American Journal of Psychology*, 114, 175–192. https://doi.org/10.2307/1423513

Khuder, B., & Harwood, N. (2015). L2 writing in test and non-test situations: Process and product. *Journal of Writing Research*, 6, 233–278. https://doi.org/10.17239/jowr-2015.06.03.2

Kim, M., Tian, Y., & Crossley, C. A. (2021) Exploring the relationships among cognitive and linguistic resources, writing processes, and written products in second language writing. *Journal of Second Language Writing*, 53, Article 100824. https://doi.org/10.1016/j.jslw.2021.100824.

Kormos, J. (2021, October 28). *Working memory: Can we make it work for inclusive language teaching?* [Conference session]. Plenary speech at the European Second Language Conference, Barcelona, Spain.

Kormos, J. (2023) The role of cognitive factors in second language writing and writing to learn a second language. *Studies in Second Language Acquisition*, 45, 622–646.

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

Lee, J. (2019). *The effects of time constraints, genre, and proficiency on L2 writing fluency behaviors and linguistic outcomes.* [Unpublished doctoral dissertation]. Michigan State University.

Lindgren, E., & Sullivan, K. P. H. (2006). Writing and the analysis of revision: An overview. In Sullivan, K. P. H. & Lindgren, E. (Eds.), *Computer key-stroke logging: Methods and applications* (pp. 31–44). Elsevier.

Lu, X. (2022). Second language Chinese computer-based writing by learners with alphabetic first languages: Writing behaviors, second language proficiency, genre, and text quality. *Language Learning*, 72, 45–86. https://doi.org/10.1111/lang.12469

Lu, X., & Révész, A. (2021). Revising in a non-alphabetic language: The multi-dimensional and dynamic nature of online revisions in Chinese as a second language. *System*, 100, 1–13. https://doi.org/10.1016/j.system.2021.102544

Manchón, R. M., & Roca de Larios, J. (2007). Writing-to learn in instructed language learning contexts. In E. Alcón Soler & M. P. Safont Jordà (Eds.), *Intercultural language use and learning* (pp. 101–121). Springer. https://doi.org/10.1007/978-1-4020-5639-0_6

Manchón, R.M., Roca de Larios, J., & Murphy, L. (2009). The temporal dimension and problem-solving nature of foreign language composing. Implications for theory. In R. M. Manchón (Ed.), *Foreign language writing: Learning, teaching and research* (pp. 102–124). Multilingual Matters.

McCormick, T., & Sanz, C. (2022). Working memory and L2 grammar learning among adults. In J. Schwieter & Z. Wen (Eds.), *The Cambridge handbook of working memory and language* (pp. 573–592). Cambridge University Press. doi:10.1017/9781108955638.032

Michel, M., Kormos, J., Brunfaut, T., & Ratajczak, M. (2019). The role of working memory in young second language learners' written performances. *Journal of Second Language Writing*, 45, 31–45. https://doi.org/10.1016/j.jslw.2019.03.002

Michel, M., Révész, A., Lu, X., Kourtali, N., Lee, M., & Borges, L. (2020). Investigating L2 writing processes across independent and integrated tasks: A mixed methods study. *Second Language Research*, 36, 277–304. https://doi.org/10.1177/0267658320915501

Miyake, A., Emerson, M. J., Padilla, F., & Ahn, J. C. (2004). Inner speech as a retrieval aid for task goals: The effects of cue type and articulatory suppression in the random task cuing paradigm. *Acta Psychologica*, 115, 123–142. https://doi.org/10.1016/j.actpsy.2003.12.004

Nicolás-Conesa, F., Roca de Larios, J., & Coyle, Y. (2014). Development of EFL students' mental models of writing and their effects on performance. *Journal of Second Language Writing*, 24, 1–19. https://doi.org/10.1016/j.jslw.2014.02.004

Plonsky, L., & Oswald, F. L. (2014). How big is "big"? Interpreting effect sizes in L2 research. *Language Learning*, 64, 878–912. https://doi.org/10.1111/lang.12079

Révész, A., Kourtali, N., & Mazgutova, D. (2017). Effects of task complexity on L2 writing behaviors and linguistic complexity. *Language Learning*, 67, 208–241. https://doi.org/10.1111/lang.12205

Révész, A, Michel, M., & Lee, M. (2017). *Investigating IELTS Academic Writing Task 2: Relationships between cognitive writing processes, text quality, and working memory* International English Language Testing System Partners. https://www.ielts.org/for-researchers/research-reports/ielts_online_rr_2017-3

Révész, A., Michel, M., & Lee, M. (2019). Exploring second language writers' pausing and revision behaviours: A mixed-methods study. *Studies in Second Language Acquisition*, 41, 605–631. https://doi.org/10.1017/S027226311900024X

Révész, A., Michel, M., Lu, X., Kourtali, N., Lee, M., & Borges, L. (2022). The relationship of proficiency to speed fluency, pausing and eye-gaze behaviours in L2 writing. *Journal of Second Language Writing*, 58, Article 100927. https://doi.org/10.1016/j.jslw.2022.100927

Rijlaarsdam, G., & Van Den Bergh, G. (1996). The dynamic of composing—An agenda for research into an interactive compensatory model of writing: Many questions, some answers. In C. M. Levy & S. Ransdell (Eds.), *The science of writing: theories, methods, individual differences, and applications* (pp. 107–126). Lawrence Erlbaum.

Roca de Larios, J., Manchón, R., Murphy, L., & Marín, J. (2008). The foreign language writer's strategic behavior in the allocation of time to writing processes. *Journal of Second Language Writing*, 17, 30–47. https://doi.org/10.1016/j.jslw.2007.08.005

Schilperoord, J. (1996). *It's about time: Temporal aspects of cognitive processes in text production*. Rodopi. https://doi.org/10.1075/sl.21.3.09har

Serafini, E., & Sanz, C. (2016). Evidence for the decreasing impact of cognitive ability on second language development as proficiency increases. *Studies in Second Language Acquisition*, 38(4), 607–646. https://doi.org/10.1017/S0272263115000327

Spelman Miller, K. (2000). Academic writers on-line: Investigating pausing in the production of text. *Language Teaching Research*, 4, 123–148. https://doi.org/10.1177/136216880000400203

Stevenson, M., Schoonen, R., & de Glopper, K. (2006). Revising in two languages: A multi-dimensional comparison of online writing revisions in L1 and FL. *Journal of Second Language Writing*, 15, 201–233.

Thorson, H. (2000). Using the computer to compare foreign- and native-language writing processes: A statistical and case study approach. *The Modern Language Journal*, 84, 55–70. https://doi.org/10.1111/0026-7902.00059

Tillema, M. (2012). *Writing in first and second language. Empirical studies on text quality and writing processes* [Unpublished doctoral thesis]. Netherlands Graduate School of Linguistics (LOT). https://www.lotpublications.nl/Documents/299_fulltext.pdf

Unsworth, N., Heitz, R.P., Schrock, J.C. & Engle, R.W. (2005). An automated version of the operation span task. *Behavior Research Methods*, *37*, 498–505. https://doi.org/10.3758/BF03192720

Vallejos, C. (2020). *Fluency, working memory and second language proficiency in multicompetent writers.* [Unpublished doctoral dissertation]. Georgetown University.

Van Weijen, D. (2009). *Writing processes, text quality, and task effects.* Empirical studies in first and second language writing, 201. Netherlands Graduate School of Linguistics.

Van Waes, L., & Leijten, M. (2015). Fluency in writing: A multidimensional perspective on writing fluency applied to L1 and L2. *Computers and Composition*, *38*, 79–95. https://doi.org/10.1016/j.compcom.2015.09.012

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

Xu, C., & Qi, Y. (2017). Analyzing pauses in computer-assisted EFL writing: A computer-keystroke-log perspective. *Journal of Educational Technology and Society*, *20*, 24–34.

Zhao, Y. (2013). Working memory and corrective recasts in L2 oral production. *Asian Journal of English Language Teaching*, *23*, 57–82.