# Research Article

## EXPLORING SECOND LANGUAGE WRITERS' PAUSING AND REVISION BEHAVIORS

A MIXED-METHODS STUDY

**Andrea Révész\***

*University College London*

**Marije Michel**

*University of Groningen and Lancaster University*

**Minjin Lee**

*Ewha Womans University*

**Abstract**

This study investigated the cognitive processes underlying pauses at different textual locations (e.g., within/between words) and various levels of revision (e.g., below word/clause). We used stimulated recall, keystroke logging, and eye-tracking methodology in combination to examine pausing and revision behaviors. Thirty advanced Chinese L2 users of English performed a version of the IELTS Academic Writing Task 2. During the writing task, participants' key strokes were logged, and their eye movements were recorded. Immediately after the writing task, 12 participants also took part in a stimulated recall interview. The results revealed that, when participants paused at larger textual units, they were more likely to look back in the text and engage in higher-order writing processes. In contrast, during pauses at lower textual units, they tended to view areas closer to the inscription point and engage in lower-order writing processes. Prior to making a revision, participants most frequently had viewed the text that they subsequently revised or their eye gazes had been off-screen. Revisions focused more on language- than content-related issues, but there was a smaller difference in the number of language- and content-focused stimulated recall comments when larger textual units were revised.

---

## INTRODUCTION

The past two decades have witnessed a growing interest in describing the online behaviors of second language (L2) writers, that is, the directly observable features of the writing process. An increasing amount of research has also been concerned with investigating the cognitive macro-writing processes (e.g., planning, translation) and subprocesses (e.g., planning content, lexical encoding) (Manchón, Roca de Larios, & Murphy, 2007) that underlie L2 writing behaviors. Among the writing behaviors studied, pausing and revision phenomena have probably received the most attention (e.g., Roca de Larios, Manchón, Murphy, & Marín, 2008; Van Waes & Leijten, 2015). This increased attention has been driven by both theoretical and practical concerns. At the theoretical front, researchers have studied pausing and revision behaviors to test models of L2 writing, presuming that characteristics of pausing and revision are reflections of the cognitive processes in which writers engage (Baaijen, Galbraith, & de Glopper, 2012). The investigation of pausing and revision phenomena is also of significance to the areas of L2 assessment and instruction. Information about the cognitive processes associated with patterns of pausing and revision may help diagnose areas of writing difficulty, aiding L2 educators in identifying gaps in students' L2 knowledge and skills and thereby tailoring instruction to meet their needs.

Besides theoretical and practical considerations, the enhanced research effort at studying pausing and revision behaviors is probably due to recent technological developments, which allow for obtaining a more fine-grained description of observable pausing and revision phenomena and, hence, for making more valid inferences about corresponding cognitive processes. For many years, verbal protocols were the preferred method in writing process research (e.g., Roca de Larios et al., 2008), but, increasingly, L2 researchers also utilize more novel tools such as keystroke logging (Spelman Miller, 2000; Stevenson, Schoonen, & de Glopper, 2006) and eye-tracking to examine pausing and revision behaviors (Chukharev-Hudilainen, Feng, Saricaoglu, & Torrance, 2019; Gánem-Gutiérrez & Gilmore, 2018; Révész, Michel, & Lee, 2017). A few studies have additionally succeeded in combining multiple techniques to gain a more complete picture of pausing and revision phenomena and underlying cognitive processes (e.g., Chukharev-Hudilainen et al., 2019; Khuder & Harwood, 2015; Révész, Kourtali, & Mazgutova, 2017; Stevenson et al., 2006).

The aim of the present study was to contribute to and expand on existing research on cognitive processes associated with pausing and revision behaviors. In particular, we intended to gain insights into the cognitive processes underlying pauses at different textual locations (e.g., within words, between sentences) and various levels of revision (e.g., below word, clause, and above). We used stimulated recall, keystroke logging, and eye-tracking methodology together to investigate pausing and revision phenomena, the primary contribution of our study being methodological in nature. In the area of L2 writing, little research exists that has employed eye-tracking to examine processes in relation to different types of pausing and revision, and, to the best of our knowledge, this study constitutes one of the first attempts to combine it with stimulated recall and keystroke logging data simultaneously. This combination of quantitative and qualitative methods allowed us, based on a single dataset, to triangulate information about L2 writers' thought processes during pauses and revisions (stimulated recall), real-time text

production behaviors (keystroke logging), as well as viewing behaviors including reading during pauses and before revisions (eye-tracking). As a consequence, we were able to obtain a fuller description and understanding of pausing and revision phenomena than could be achieved in previous studies.

## LITERATURE REVIEW

### THE SECOND LANGUAGE WRITING PROCESS

We used Kellogg's (1996) model of writing as the theoretical basis for this investigation. Our rationale for adopting this model to frame the study was that, compared to other models of writing (e.g., Galbraith, 2009; Hayes, 2012), this framework puts greater emphasis on the linguistic encoding processes involved in transforming the writer's intended content into text. These processes are expected to pose considerable difficulty for L2 writers for whom text generation, including lexical retrieval, syntactic encoding, and expression of cohesion, is more effortful and less automatic than for L1 users whose linguistic encoding skills tend to be more automatized (Kormos, 2012; Roca de Larios, Murphy, & Manchón, 1999).

Kellogg conceptualizes writing as an interactive and cyclical process, which entails the subprocesses of formulation, execution, and monitoring. At the formulation stage, writers plan the content of the written text and translate it into linguistic code. While they plan, writers are involved in higher-order writing processes such as retrieving ideas from their long-term memory and/or the task input, and arranging these to produce a coherent plan for what to include in the written text and how to organize the content. In the course of translation, the writer translates the content planned into linguistic form through engaging in lower-order writing processes, including lexical retrieval, syntactic encoding, and use of cohesive devices. During the execution stage, writers employ motor movements to create a typed or handwritten piece. Finally, in the monitoring phase, the writer checks whether the text appropriately expresses the content they planned. If discrepancies are identified, then revisions may follow to ensure that the text is an appropriate expression of the writer's plan.

To assess this and other cognitive models of writing (e.g., Bereiter & Scardamalia, 1987; Flower & Hayes, 1980; Hayes, 1996), researchers have often turned to studying pausing and revision behaviors, assuming that pauses are observable correlates of underlying cognitive processes in general and the type of revisions made can give insights into the nature of monitoring in particular. In the sections that follow, we provide a review of previous research exploring writing processes through the study of pausing and revision behaviors, with a particular emphasis on the methodological aspects of earlier research.

### PAUSING BEHAVIORS AND UNDERLYING COGNITIVE PROCESSES

Pausing, defined here as the absence of typing or handwriting, may be the manifestation of a variety of underlying writing processes. Pauses may reflect cognitive activities (e.g., planning, linguistic encoding, rereading previously produced text), but may also occur due to physical (e.g., executing motor movements while typing or handwriting)

and sociopsychological (e.g., daydreaming) factors (Alves, Castro, de Sousa, & Stromqvist, 2007; Wengelin, 2006). Although inferring the exact reason(s) for pausing is challenging, it appears that, depending on where and how long writers pause, pauses are likely to signal differential underlying processes. Researchers have put forward two specific assumptions regarding the relationships between cognitive activities and the location and frequency of pauses. First, pausing at higher-level textual units (e.g., between clauses and sentences) is more likely to reflect higher-order writing subprocesses such as planning content and organization, whereas pauses at lower textual units (e.g., within and between words) tend to be associated with lower-level writing subprocesses, including the retrieval of lexical items and encoding of morphology (Schilperoord, 1996). Second, length of pausing before a textual unit has been argued to reflect the mental effort involved in the planning and translation processes associated with the production of the forthcoming textual unit (e.g., Damian & Stadthagen-Gonzalez, 2009). Taken together, pauses between higher textual units are expected to be longer than pauses within and between lower textual units, given that the assembly of higher textual units is anticipated to demand more cognitive effort.

These assumptions are consistent with the findings of a number of L1 empirical studies involving both children and adults. For example, Chanquoy, Foulin, and Fayol (1990), in a carefully designed experimental study, asked children and adults to write endings for orally presented texts. The endings that participants had to produce differed in terms of predictability (trivial or unexpected ending required) or syntactic complexity (one or several sentences needed). The researchers found that, when participants were asked to write predictable or less syntactically complex endings, they displayed shorter prewriting pause durations. This was interpreted as reflecting the reduced cognitive load involved in planning the forthcoming text. In a more recent study, van Hell, Verhoeven, and van Beijsterveldt (2008) studied the pausing behaviors of children and adults while composing narrative or expository texts using a digitiser tablet to record handwriting movements. Similar to Chanquoy et al. (1990), a key finding of the study was that both children and adults displayed longer pauses at boundaries between higher textual units, suggesting that the writers spent more time planning and/or formulating their next idea. Parallel trends were reported in several studies of L1 writing, which investigated the pausing behaviors of adult writers using keystroke logging methodology (e.g., Medimorec & Risko, 2017; Van Waes & Leijten, 2015; Van Waes & Schellens, 2003).

In assessing whether similar patterns apply in L2 writing, most researchers have also relied on keystroke logging methodology, that is, recording the writers' keystrokes and mouse movements while writing. Spelman Miller (2000) was one of the first studies to compare length of pausing across different textual locations in L2 writing. The participants, 10 L1 and 11 L2 writers of English, wrote an evaluative and a descriptive essay while their online keystrokes and mouse movements were recorded. The resulting log files were analyzed in terms of several fluency and pausing measures. In line with patterns emerging from L1 writing research, Spelman Miller found that length of pausing increased with increasing level of textual units, with the longest pauses occurring between sentences, followed by pauses between clauses, intermediate constituents, and words, and within words. The same pattern was observed for the two task types and for the two groups of writers (L1 vs. L2), although the L2 writers, as expected, generally paused longer at each textual location.

Spelman Miller's findings have been confirmed in a number of more recent studies employing keystroke logging (Chukharev-Hudilainen et al., 2019; Révész, Kourtali et al., 2017; Révész, Michel et al., 2017; Van Waes & Leijten, 2015). Among these, Van Waes and Leijten's work is of particular significance because the researchers used four different pause thresholds (200, 500, 1000, 2000ms) when studying L2 fluency behaviors. The participants were 68 university students, who wrote two descriptive texts, one in their L1 (Dutch) and one in their L2 (English, French, Spanish, or German). For both populations, Van Waes and Leijten observed, like Spelman Miller, that, as textual units increased, the length of the pauses preceding the textual units increased. Importantly, this trend was maintained for all four pause thresholds. To sum up, assuming that longer pauses are indeed a reflection of greater mental effort, the overall results of keystroke logging studies indicate that L2 writers, similar to their L1 counterparts, find it more cognitively demanding to produce longer stretches of text.

The sole use of keystroke logs, however, does not allow for making inferences about the specific cognitive processes that underlie pausing behaviors. Pauses of similar lengths may reflect various cognitive activities, such as planning content, difficulty with translation, rereading of previous text, and/or revision of planned language in the form of inner speech (Baaijen et al., 2012). A possible way to obtain more detailed information about the cognitive processes that underlie pausing at various textual units is to combine keystroke logging methodology with other techniques such as verbal reports and eye-tracking. Eye-tracking allows for the recording of writer's moment-to-moment eye-gaze behaviors during writing, thus it can capture viewing processes such as the rereading of instruction or previously produced text during pauses. However, a remaining limitation of the joint use of keystroke-logging and eye-tracking data is that it can provide no direct evidence into the cognitive processes of L2 writers while they pause. Combining these techniques with verbal protocols can help resolve this issue. Verbal reports can shed light on the purpose of reading, whether it is to monitor performance or to generate new ideas. In addition, verbal reports can provide insights into writers' conscious cognitive activities when their eye fixations are off-screen; for example, whether they engage in planning content, linguistic encoding, and/or inner speech.

Although there is a growing number of studies utilizing a combination of methods to tap the writing process (e.g., Gánem-Gutiérrez & Gilmore, 2018; Khuder & Harwood, 2015; Révész, Kourtali et al., 2017; Stevenson et al., 2006), only few such L2 studies (Chukharev-Hudilainen et al., 2019; Révész, Kourtali et al., 2017) have looked into pausing behaviors according to textual location. Révész, Kourtali et al. (2017) studied the writing behaviors of 73 advanced L2 writers carrying out tasks of differential cognitive complexity. In addition to recording the participants' online writing behaviors by keystroke logging software, the researchers invited eight participants to describe their thought processes using stimulated recall, elicited by the playback of their keystroke recordings. As mentioned previously, the results for pause length patterned with other studies, with longer pauses occurring between higher textual units, regardless of whether participants engaged in cognitively simple or complex task performance. The only exception to this trend was similar pause lengths observed for pauses between words and clauses. The stimulated recall comments further revealed that, parallel to what was proposed for L1 writing (Schilperoord, 1996), pausing at higher textual units was more likely to be linked to higher-level writing processes. When recalling their thoughts during

between-sentence pauses, participants referred to planning-related processes considerably more frequently irrespective of task complexity condition. Révész and colleagues concluded that, indeed, longer pausing, which was observed before the production of larger textual units, tended to reflect engagement in higher-order writing processes.

Instead of utilizing verbal protocols, Chukharev-Hudilainen et al. (2019) combined keystroke logging with eye-tracking to study L2 writing fluency. The participants were 24 L1 speakers of Turkish, who composed two argumentative essays, one in Turkish and one in L2 English. The keystroke logs yielded longer pauses between larger textual units, similar to the overall trend observed in Révész, Kourtali et al. (2017). One exception to this pattern was the similar pause lengths found preceding words and nonfinite clauses in L2 writing, a finding also consistent with Révész, Kourtali et al.'s (2017) results (although this study did not code for different clause types). The eye-gaze data revealed that, overall, writers were more likely to view their previously produced text before the formulation of larger linguistic units. Interestingly, however, the likelihood of looking back, for L2 writers, was lower at the start of finite clauses as compared to pauses before other textual units. The study also found that lookback distances were similar prior to subsentence units, but participants had gone significantly further back in their text before they composed a new sentence. Taken together, the findings of Chukharev-Hudilainen et al. indicate that longer pauses preceding higher textual units are associated, at least in part, with rereading longer stretches of previously produced texts.

Although Révész, Kourtali et al. (2017) and Chukharev-Hudilainen et al. (2019) provide more detailed accounts of the processes underlying pausing behaviors than studies that have used keystroke logging alone, they are not without shortcomings. Révész, Kourtali et al. (2017) sheds little light on participants' viewing behaviors during pauses, whereas Chukharev-Hudilainen et al. (2019) provides no direct information about participants' thought processes while writing. To address these limitations, the present study made use of all three data sources—keystroke logging, eye-tracking, and verbal protocol—to better uncover the cognitive processes associated with pausing at various textual units.

### REVISION BEHAVIORS AND UNDERLYING COGNITIVE PROCESSES

Revision constitutes a complex set of cognitive activities, involving the subprocesses of reading, evaluating, and changing previously produced text and revising planned and/or translated ideas internally before they are physically transcribed into text (e.g., Broekkamp & van den Bergh, 1996; Stevenson et al., 2006). Revisions may be concerned with various aspects of writing. Writers may alter the meaning or the information conveyed in the text; they may modify the grammar or lexis used to express the intended content without changing the core information; or they may revise because they have committed graphic or typographic errors (Stevenson et al., 2006).

Several taxonomies have been put forward to model different types of revision processes and outcomes (Faigley & Witte, 1981; Lindgren & Sullivan, 2006a, 2006b; Matsuhashi, 1987; Porte, 1996, 1997; Roca de Larios et al., 1999; Scardamalia & Bereiter, 1987; Stevenson et al., 2006; Thorson, 2000). Of these, the frameworks proposed by Lindgren and Sullivan (2006a, 2006b) and Stevenson et al. (2006) are the most comprehensive, proposing a similar hierarchical structure of categories. Lindgren

and Sullivan distinguish between internal and external revisions, the former taking place in the writer's head (possibly manifest in pausing behaviors) and the latter entailing visible alterations to the text. External revisions may be further subdivided into pre-contextual and contextual revisions. Precontextual revisions occur at the point of inscription; in other words, there is text produced before, but not after, them. Contextual revisions are carried out away from the point of inscription; that is, they occur in context, preceded and followed by previously written text. Both precontextual and contextual revisions may alter conceptual (e.g., ideas) or form-related (e.g., grammar) aspects of the text. Our study investigated the processes underlying external revisions, both contextual and precontextual. An in-depth study of internal revisions was beyond the scope of this article.

A large part of L2 research on revision has been concerned with exploring what factors may influence the type of revisions in which L2 writers engage. Earlier work has observed that, in general, writers with lower proficiency are more likely to focus on linguistic, lower-level aspects of their text during revision (e.g., Barkaoui, 2016; Porte, 1996; Whalen & Ménard, 1995). Probably due to their limited and less automatized L2 knowledge, low-proficiency writers experience greater cognitive load when revising language-related issues, resulting in fewer attentional resources left for higher-order revision processes (e.g., reusing ideas) (Broekkamp & van den Bergh, 1996). The cognitive complexity of the writing task has also been found to influence the type of revision that L2 writers carry out. Révész, Kourtali et al. (2017), in addition to pausing, also looked into the effects of task complexity on revision processes, and found that more conceptually demanding tasks led to fewer revisions below the word level. The authors interpreted this finding as suggesting that, owing to the greater cognitive demands posed by the task, writers might have had less attention left to allocate to lower-level revisions (see, however, Thorson, 2000). Besides proficiency and task complexity, contextual variables such as writing under test versus nontest conditions (Khuder & Harwood, 2015) or producing typed versus handwritten texts (Li, 2006) have also been shown to affect the type of revision processes in which L2 writers are involved.

Turning to methodological issues, researchers have relied on a variety of techniques to tap L2 revision behaviors, including verbal protocols such as the think-aloud procedure (Roca de Larios et al., 2008; Whalen & Ménard, 1995), video recordings (Matsuhashi, 1981), keystroke logging (Barkaoui, 2016; Thorson, 2000), and screen-capture programs (Elola & Mikulski, 2013). Like studies of pausing, experiments investigating revision behaviors are also beginning to utilize elicitation methods in combination to compensate for the limitations associated with the use of individual techniques. Stevenson et al. (2006) were among the first to employ keystroke logging together with the think-aloud procedure to investigate type of revisions made by L2 writers. The aim of the study was to test the hypothesis that, when students compose in their L2 rather than their L1, attention to linguistic processes may inhibit higher-level conceptual processing. The participants were 22 Dutch junior high school students, who composed a text in both Dutch and L2 English. The researchers found little evidence for the assumption that higher-order writing processes are constrained in L2 writing. Khuder and Harwood (2015) and Révész, Kourtali et al. (2017), two studies mentioned earlier, also used a combination of methods (keystroke logging, stimulated recall, and screen-capture

software) to gain information about the type of revision processes in which writers engaged.

The joint application of methods in these studies, just as in research on pausing, allowed researchers to arrive at more valid and fine-tuned conclusions about revision processes. However, existing research provides little information about viewing behaviors in relation to revision. Given that rereading and evaluation are key revision subprocesses, it would appear fruitful to elicit eye-gaze recordings while students compose a text and triangulate these with other data sources. For example, eye-tracking enables researchers to obtain direct evidence about what parts of the texts and/or instruction participants have viewed prior to making a revision. To exploit the affordances of this technique, we adopted a mixed-methods design to study revision behaviors, employing eye-tracking together with keystroke logging and stimulated recall. It was hoped that by gaining information about writers' conscious cognitive activities during revision through stimulated recall, and capturing their real-time revision behaviors, conscious or unconscious, through keystroke logging and eye-tracking will aid in obtaining a comprehensive account of revision behaviors and associated cognitive processes.

## RESEARCH QUESTIONS

We formulated the following research questions:

1.  What are the cognitive processes underlying the pausing behaviors of L2 writers on an academic essay task, as reflected in
    a.  participants' eye-gaze behaviors during pauses at different locations?
    b.  stimulated recall comments associated with different pause locations?
2.  What are the cognitive processes underlying the revision behaviors of L2 writers on an academic essay task, as reflected in
    a.  participants' eye-gaze behaviors before revisions at different levels?
    b.  stimulated recall comments associated with revisions at different levels?

In the present study, pause location was operationalized in terms of whether participants paused within a word, between words, or between sentences. Level of revision was defined based on whether the revision concerned a change below the word level, at the word level, below the clause level, at the clause level or above, or at the sentence level and above. Participants' eye-gaze behaviors were categorized according to the level of the textual unit (e.g., word, phrase, sentence) that had been viewed during the pause or immediately before the revision.

## METHOD

### DESIGN

The dataset for the present study was collected as part of a larger project investigating the relationships between cognitive writing processes, text quality, and working memory capacity reported in Révész, Michel et al. (2017). The current study delves into a more in-depth analysis of pausing and revision phenomena by examining the eye-gaze behaviors

and stimulated recall comments of participants in relation to pause location and level of revision. With this aim in mind, we analyzed the writing performances of 30 L2 writers on a version of Task 2 of the IELTS Academic Writing Test. The participants' online writing behaviors were captured with the keystroke logging software Inputlog 6.1.5 (Leijten & Van Waes, 2013) and a Tobii X2-60 mobile eye-tracking system. Twelve participants were additionally invited to take part in a stimulated recall session. Thus, the study adopted a mixed-methods design, allowing for the triangulation of quantitative and qualitative data sources.

### PARTICIPANTS

All 30 participants were L2 users of English with Mandarin as their first language. They were all international students at a university in the United Kingdom, and had an overall score of 7 or higher on the IELTS test, equivalent to C1 or higher in the Common European Framework of Reference (CEFR). The majority were female ($n = 27$), and their age ranged from 18 to 34 with a mean of 26.60 ($SD = 3.69$). Most of the participants were studying toward a masters' level degree ($n = 24$), five students were working on a doctorate, and one participant was enrolled in a bachelor's course. The third author who conducted the data-collection sessions was not acquainted with the participants; she met them through the data-collection session.

### INSTRUMENTS AND PROCEDURES

#### Writing task

A computer-based version of Task 2 of the IELTS Academic Writing Test was used as an elicitation instrument. The essay prompt that the participants were asked to address was:

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.

Participants had no planning time and received 40 min to complete the task. On average they spent 34 min (SD = 7 min 14 sec) on task completion. They wrote in an Microsoft Word document, which was set to the monospace font type Consolas with font size 16 and 1.5 point spacing between lines to allow for more precise eye-gaze measurement.

#### Stimulated recall

The aim of the stimulated recall sessions was to elicit the thought processes in which participants ($n = 12$) engaged when carrying out the IELTS writing task. The participants' recall was prompted by a screen replay of their keystrokes and eye movements during their writing performance. They were told in everyday language that the red

circles (eye fixations) and lines (saccades) in the recordings indicated their eye movements, and that larger circles meant that they had fixated longer. They were also encouraged to pause the recording at any point they wished to describe the thoughts they had during the writing task. The researcher additionally stopped the recording when participants paused, made a revision, went back to parts of the text they had written earlier, or produced unusual or interesting eye movements (e.g., longer fixations, regressions) but did not comment on these behaviors on their own. It was emphasized that participants should only report what they were thinking at the time they carried out the task. The stimulated recall sessions were conducted in English. Given the high proficiency level of the participants, this did not seem to cause difficulty. The stimulated recall sessions were video-recorded to capture not only participants' verbal comments but also spatial movements (e.g., pointing to the screen). The sessions lasted between 60 and 90 min.

## DATA COLLECTION

All the participants took part in one individual session in the first author's office. After giving informed consent, they were administered a short background questionnaire. This was followed by the calibration of the eye-tracker, a mobile Tobii X2-60 with a temporal resolution of 60 Hz. The eye-tracker was mounted to a 23-inch screen, with the participants seated about 60 cms away from the center of the screen. A 9-point calibration grid was used, and the experiment was presented with Tobii Studio 3.0.9 software (Tobii Technology, n.d.). After the eye-tracker had been calibrated, participants were asked to complete the IELTS writing task. This was followed by the typing test. After a short break, the 12 stimulated recall participants were introduced to the stimulated recall procedure, and then invited to describe their thoughts while writing the IELTS essay based on the replay of the recording of their writing session.

## DATA ANALYSIS

### Analysis of keystroke logs

To identify pauses in the keystroke logs, we ran a pause summary analysis for each participant using Inputlog. We adopted a pause threshold of 2 s following conventions in writing research (e.g., Wengelin, 2006; see, however, Van Waes & Leijten, 2015). With the help of Inputlog, we categorized pauses according to the textual unit where they occurred, whether they were located within words, between words, or between sentences. Between-word pauses were treated as one pause, given that pauses between words often include one pause before the spacebar is pressed and one pause before the beginning of the next word. We also extracted measures of pause frequency and pause length by location (the results for these indices are also reported in Révész, Michel et al., 2017).

We also employed the Inputlog software to identify revisions. Then, we manually coded revisions in terms of whether they involved a change below the word level (i.e., one or more characters but less than a whole word), at the word level (i.e., a whole word), below the clause level (more than a word but less than a clause), at the clause level

and above (one clause or more but less than a sentence), or at the sentence level and above (one sentence or more). Ten percent of the data was randomly selected and coded by a second researcher. Cohen's kappa was found to be .96 (SE = .01) based on 318 decisions, that is, intercoder agreement was high.

### Analysis of eye-tracking data

To gain further insights into the nature of participants' online writing behaviors, we reviewed participants' eye-gaze behaviors during pausing and before revisions. First, we searched for all pauses (threshold: 2 s) and revisions in the Inputlog files, and then viewed the eye-gaze recordings with the help of Tobii Studio 3.0.9 software to identify the same points in time in the eye-gaze data. Once the pauses and revisions in the Inputlog files and eye-gaze recordings had been matched, participants' eye movements were qualitatively categorized by visually inspecting the eye-gaze recordings using the pauses and revisions identified in the Inputlog files as reference points.

For all pauses, participants' eye movements were coded in terms of whether their eye gaze(s) remained during the pause at the point of inscription or visited areas within the word/phrase, clause, sentence, or paragraph preceding the point of inscription. Given the qualitative nature of this coding procedure, we did not consider number of fixations, we only coded for the presence/absence of fixation(s) within a specific area during a pause. In cases in which participants visited several textual units during a pause, the largest textual unit visited was used as the code for the pause. For example, when a participant fixated on a point/points both within and outside the preceding clause but within the preceding sentence, this series of fixations was coded as "sentence." To illustrate this, Figure 1 shows two screen shots of text production with overlaying eye gazes (circles). At the top of both pictures, the task prompt is visible in slightly smaller font size. The larger writing pane on the left shows a participant pausing after having written "because." The eye gazes reveal viewing within the preceding sentence starting with "Such a…," which was coded as "sentence." On the right, the writer stopped after having written "I." The eye gazes reveal viewing behavior around that word but also beyond the sentence boundary focusing on the earlier sentence starting with "Studying abroad…," which was coded as "paragraph."
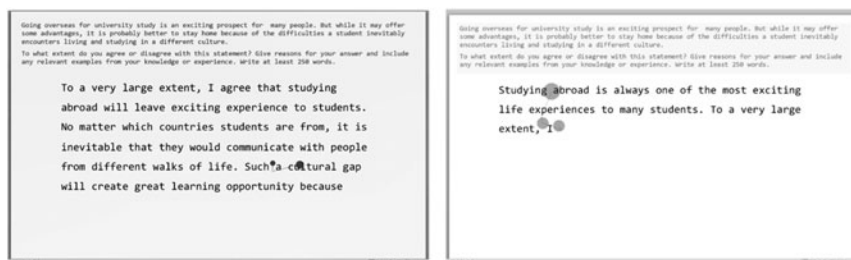


FIGURE 1.    Examples of scanpaths for eye gazes during pauses at sentence (left) versus paragraph (right) level. On the left the eye fixations (indicated by circles) stay within the sentence preceding the inscription point, whereas on the right one of the eye fixations is beyond the sentence preceding the inscription point.

For revision, we considered viewing behaviors before the revision, whether participants fixated on area(s) within the word/phrase, the clause, the sentence, or the paragraph before the point of inscription. Similar to pausing, we did not code for number of fixations within areas; we exclusively focused on whether a fixation occurred within an area or not before a revision. For each revision, the code was specified as the largest textual unit participants gazed at before the revision. To give an example, when a participant fixated on an area/areas in the previous word/phrase and beyond but within the preceding clause, this fixation/fixations was coded as "clause." Occasionally, participants went back to the instructions or did not view the computer screen while they paused or before they revised. These instances were coded as "instruction" and "off-screen," respectively. Ten percent of the pausing and revision data, randomly selected, were double-coded by one of the researchers. Cohen's kappa was found to be very good (n = 654, Kappa: .90, SE = .02).

To control for differences in pause/revision frequency across participants, we divided the counts for each participant for each textual unit by the number of times they paused/revised (overall and at various pause locations/levels of revision). We used the resulting proportions in further analyses.

### Analysis of stimulated recall comments

The stimulated recall data comprised 547 min, with an average of 46 min and 35 s per participant. The analysis of the comments involved five steps. First, the data were transcribed. Second, the first and third author independently reviewed the pause- and revision-related comments and identified emergent categories. Third, the resulting micro-categories were grouped into more general categories informed by Kellogg's (1996) model of writing. These general categories and examples for them are presented in Tables 1 and 2 for pausing and revision, respectively. Intercoder percentage agreement for category identification was found to be high (96%), and discrepancies between the researchers were resolved through discussion. Fourth, the third author coded all the comments by annotating the data based on the agreed coding scheme. To check intercoder agreement, the first author also coded the data for three participants, randomly selected. The agreement between the first and second coder reached a good level (n = 85, Kappa: .77, SE = .05). Finally, to form a frequency count for each participant, the comments falling into specific categories were added up.

### Statistical analyses

A series of nonparametric Friedman tests of differences among repeated measures was computed to test whether there were differences in the frequency with which participants viewed various levels of textual units at different pause locations and before different levels of revision. When the overall Friedman test was found significant, follow-up Wilcoxon Signed Rank tests were computed to identify pairwise differences. The alpha level was set at .05 for all tests, given the relatively small sample size. Effect size values were calculated using the formula $r = Z/sqrt(N)$. Following Plonsky and Oswald (2014), values larger than .25, .40, and .60 were considered as small, medium, and large, respectively.

TABLE 1. Examples for stimulated recall comments: Pausing

| Process/Subprocess | Example |
|---|---|
| Planning | Do I agree or disagree? Which position should I take? Which |
|   Content |   one is easy to write? Which side is easier to take? |
|   Organization | At that time, I was keeping on the eye on the word count. I found my word count is almost 250. I didn't have much space to develop my argumentation too much. I remembered that I wrote "first of all," then... there should be "secondly" or "furthermore." I realized that maybe I have space for only one opinion in detail. |
| Formulation | Because I've already used the word "discussions" so I was |
|   Lexical Retrieval |   trying to think of another word which has the same meaning. |
|   Syntactic Encoding | I was thinking whether I should treat "study abroad" as a singular or plural form. |
|   Cohesion | I was thinking about linking words I should use. "Secondly" is boring one. Should I use that? |
| Unspecified | How to say. I mean very often I can figure out how to write smoothly in a simple way. I read lots of papers and I was greatly impacted by their way of expressing. I was trying to say a sentence a little bit in a complicated way…. So it looks professional and academic. |
| Monitoring | I review from the beginning, checking any grammar mistakes. I am proofreading. |

## RESULTS

### EYE-GAZE BEHAVIORS AT DIFFERENT PAUSE LOCATIONS

Table 3 provides the median percentage of eye-gaze behaviors by pause location, that is, the values in the table present the median for how many times participants' eye gazes stayed within a particular area of interest (e.g., point of inscription, previous word/ phrase) during a pause out of all the pauses they made at that location type (e.g., within words).

As Table 3 indicates, when participants paused within words, their eye gazes remained within the previous word/phrase, clause, or sentence with similar frequency; viewed area(s) in the previous paragraph and instructions slightly fewer times; and spent the least time at the point of inscription. Most frequently, however, participants' eye gazes were not detected on the screen. A Friedman test found no significant difference in the frequency with which participants viewed various levels of textual units (word/phrase, clause, sentence, or paragraph) during within-word pauses: $\chi^2$ (3, $N = 30$) = 5.19, $p = .16$.

Participants' eye movements yielded different patterns for pauses between words. Participants stayed within the previous clause most frequently, followed by views within the preceding word/phrase, paragraph, instructions, and sentence. Similar to what was observed for within-word pauses, participants' eye gazes remained least often at the point of inscription, and were most frequently found to be off-screen. A Friedman test

TABLE 2.   Examples for stimulated recall comments: Revision

| Process/Subprocess | Example |
|---|---|
| Planning | |
|   Content | I know I wanted to write a personal case of myself. So I wanted to start a sentence to bring my case to the essay. But later, you can see I regret afterwards. I deleted it. |
|   Organization | I realized I type like I'm doing free writing. According to instruction it's like IELTS writing task so I suddenly remembered because I didn't take IELTS test before but I remember there must be some … may be some kind of structure I have to follow for that kind of formal writing so I was thinking whether the way I am writing would not meet that kind of format required for the test so I thought for a while and so I stopped and changed and deleted something. |
| Formulation | I didn't want to use "competitiveness" or "competence" |
|   Lexical Retrieval | because I used them before. I chose another word "capacity." |
|   Syntactic Encoding | Because when I wrote this sentence, I didn't notice the tense and I examined it again and put the past tense. |
|   Cohesion | First I used while because I wanted to compare in the UK where I am forced to be independent and in China where I used to depend on parent and friends. First I used while but finally but is a better connection word so I used but. |
| Unspecified | I just I tried to rephrase the sentence to make it more academic. |

confirmed a significant overall difference ($\chi^2$ (3, $N = 30$) = 13.39, $p < .01$) in the median number of times participants viewed various textual units (word/phrase, clause, sentence, or paragraph). A series of follow-up pairwise Wilcoxon Signed Rank tests revealed that, when participants paused between words, they significantly less often stayed within the preceding word/phrase than the previous clause ($Z = 2.00$, $p = .04$, $r = .37$), more frequently remained within the previous word/phrase ($Z = 2.10$, $p = .04$, $r = .38$) and clause ($Z = 3.83$, $p < .01$, $r = .70$) than visiting more distant parts of the sentence. They also viewed areas in the previous paragraph significantly more often than parts of the sentence outside the previous clause ($Z = 2.04$, $p = .04$, $r = .37$). The effect sizes for these differences were close to medium or large.

  Turning to eye-gaze behaviors during pauses between sentences, Table 3 shows that, when they paused between sentences, the majority of participants did not stay at the point of inscription or within the previous word/phrase and clause, or view the instructions. They most often visited parts of the sentence beyond the preceding clause, followed by views outside the previous sentence within the paragraph. Participants' eye-gaze behaviors were observed off-screen fewer times than during within-word and between-word pauses. A Friedman test found a significant overall effect for textual location: $\chi^2$ (3, $N = 29$) = 10.00, $p = .02$. Post-hoc pairwise Wilcoxon Signed Rank tests revealed that, when participants paused between sentences, they significantly more often looked beyond the previous clause within the sentence than stayed within the previous word/phrase ($Z = 2.06$, $p = .04$, $r = .38$) and clause ($Z = 2.45$, $p = .01$, $r = .45$), and

TABLE 3.    Median percentage of eye-gaze behaviors by location of pauses and eye movements

| Pause location | N[a] | Total number of pauses Median | Location of eye movements | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | Point of inscription Median % IQ Range | Word/phrase Median % IQ Range | Clause Median % IQ Range | Sentence Median % IQ Range | Paragraph Median % IQ Range | Instruction Median % IQ Range | Off-screen Median % IQ Range | Elsewhere Median % IQ Range |
| Within words | 30 | 41 | 2% | 12% | 14% | 13% | 10% | 7.5% | 25.5% | 8% |
| | | | 5% | 18% | 12% | 9% | 10% | 8% | 26% | 11% |
| Between words | 30 | 83 | 4% | 11.5% | 18% | 8% | 10% | 9% | 23% | 4.5% |
| | | | 5% | 13% | 16% | 9% | 14% | 6% | 22% | 9% |
| Between sentences | 29 | 6 | 0% | 0% | 0% | 20% | 6% | 0% | 14% | 0% |
| | | | 0% | 17% | 17% | 50% | 20% | 14% | 50% | 6% |

[a]Sample size for categories is lower than 30 when not all participants paused at that location.

more frequently stayed within the sentence than visited areas outside the sentence in the paragraph ($Z = 2.80$, $p < .01$, $r = .52$). The effect sizes were close to or in the medium range.

Table 4 summarizes the significant patterns observed for eye-gaze behaviors during pauses.

### STIMULATED RECALL COMMENTS ASSOCIATED WITH DIFFERENT PAUSE LOCATIONS

Table 5 provides a summary of the stimulated recall comments, which were elicited to obtain insights into the cognitive processes underlying participants' pausing behavior at various pause location. Overall, the largest percentage of stimulated recall comments referred to translation processes (48%), followed by comments focusing on planning (35%) and monitoring (11%). The distribution of stimulated recall comments showed similar trends for pauses within words and between words, although the number of comments for within-word pauses was small ($n = 7$). More comments described translation (within words: 3%; between words: 38%) than planning processes (within words: 0%; between words: 23%), and comments concerning monitoring were few (within words: 0%; between words: 3%). The results for pauses between sentences, however, revealed different patterns, with a higher number of comments referring to planning as compared to translation processes.

Turning to subprocesses, in total, most of the planning comments mentioned planning content (84%), and the majority of translation comments concerned lexical encoding mechanisms (68%). The distributions were similar across pause locations for translation subprocesses. The only exception to this trend was that, for the small number of within-word pauses ($n = 6$), there was a lack of difference between the number of lexical and syntactic encoding-related comments.

To sum up, the stimulated recall data revealed that, when participants paused between sentences, they were more often concerned with planning. However, when they paused at lower textual units (within and between words), they focused on translation with greater frequency. The individual-level data for most participants also reflect these patterns.

### EYE-GAZE BEHAVIORS AT DIFFERENT LEVELS OF REVISION

Table 6 gives the median percentage of eye-gaze behaviors by level of revision, that is, the values in the table provide the median for how many times participants' eye gazes

TABLE 4.   Summary of significant patterns for eye-gaze behaviors during pauses

| | |
|---|---|
| Within words | n/a |
| Between words | clause > word/phrase, sentence; word/phrase > sentence; paragraph > sentence |
| Between sentences | sentence > word/phrase, clause, paragraph |

*Note:* > indicates a significantly larger number of views at a certain textual unit.

TABLE 5. Reasons for pausing: Number of stimulated recall comments by pause location

| Pause location | Planning | | | Translation | | | | Monitoring | No recall | Total[b] |
|---|---|---|---|---|---|---|---|---|---|---|
| | Content | Organization | All[a] | Lexical retrieval | Syntactic encoding | Cohesion | All[a] | | | |
| Within words | 1 | 0 | 1 (0%) | 3 | 3 | 0 | 6 (3%) | 0 (0%) | 0 (0%) | 7 (3%) |
| Between words | 42 | 7 | 49 (23%) | 59 | 22 | 2 | 83 (38%) | 6 (3%) | 8 (4%) | 146 (68%) |
| Between sentences | 20 | 5 | 25 (12%) | 9 | 2 | 4 | 15 (7%) | 17 (8%) | 6 (3%) | 63 (29%) |
| Total | 63 | 12 | 75 (35%) | 71 | 27 | 6 | 104 (48%) | 23 (11%) | 14 (6%) | 216 (100%) |

[a]Values for subcategories do not necessarily add up to the total, given that some comments were not specific enough to allow for further subcategorization.
[b]Due to rounding some totals do not add up to 100.

TABLE 6.    Median percentage of eye-gaze behaviors by level of revision

| Revision level | N[a] | Total number of revisions Median | Location of eye movements | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | Point of inscription Median % IQ Range | Word/phrase Median % IQ Range | Clause Median % IQ Range | Sentence Median % IQ Range | Paragraph Median % IQ Range | Instruction Median % IQ Range | Off screen Median % IQ Range | Elsewhere Median % IQ Range |
| Below word | 30 | 71 | 3% | 38% | 1% | 2% | 3% | 0% | 44% | 0% |
| | | | 7% | 26% | 3% | 4% | 10% | 0% | 42% | 2% |
| Single word | 30 | 28.5 | 0% | 34.5% | 3.5% | 7% | 7% | 0% | 34% | 0% |
| | | | 0% | 27% | 7% | 16% | 16% | 3% | 38% | 2% |
| Below clause | 30 | 24.5 | 6.5% | 29.5% | 5% | 9% | 11.5% | 0% | 29.5% | 0% |
| | | | 14% | 22% | 10% | 12% | 16% | 3% | 32% | 0% |
| Clause and above | 24 | 2 | 0% | 0% | 0% | 0% | 0% | 0% | 0% | 0% |
| | | | 13% | 0% | 30% | 33% | 50% | 0% | 48% | 0% |
| Sentence and above | 18 | 1 | 0% | 0% | 0% | 0% | 58.5% | 0% | 0% | 0% |

[a]Sample size for categories is lower than 30 when not all participants made revision at that level.

remained within an interest area (e.g., point of inscription, previous word/phrase) before making a revision out of all the revisions at that level (e.g., below word).

Table 6 shows that, when participants revised below the word level, their eye gazes stayed within the previous word/phrase considerably more frequently than the previous clause, sentence, or paragraph; remained at the point of inscription on few occasions; and were most often located off-screen. A Friedman test confirmed a significant difference for location of eye movements prior to below word-level revisions, $\chi^2$ (3, $N = 30$) = 58.84, $p < .01$. Post-hoc Wilcoxon Signed Rank tests found that this overall effect was due to significantly more instances where the eye fixations stayed within the previous word/phrase rather than visiting areas beyond the word/phrase within the previous clause ($Z = 4.78$, $p < .01$, $r = .87$), outside the clause in the sentence ($Z = 4.78$, $p < .01$, $r = .87$), and beyond the sentence in the paragraph ($Z = 4.56$, $p < .01$, $r = .83$), and to more visits to text in the preceding paragraph than in the previous clause ($Z = 3.31$, $p < .01$, $r = .60$). The effect sizes for all these relationships were large.

Similar results were obtained for revisions at the word level. Before participants revised a full word, their eye gazes most often remained within the previous word/phrase; they visited areas within the previous sentence and paragraph with considerably lower frequency; and the preceding clause had the least views. A large number of word-level revisions were preceded by eye gazes off-screen. A Friedman test identified a significant effect for eye-gaze location, $\chi^2$ (3, $N = 30$) = 49.80, $p < .01$. As a series of follow-up Wilcoxon Signed Rank tests revealed, participants remained in the previous word/phrase significantly more often than looked further in the previous clause ($Z = 4.62$, $p < .01$, $r = .84$), sentence ($Z = 4.62$, $p < .01$, $r = .84$), and paragraph ($Z = 3.86$, $p < .01$, $r = .70$). The effect sizes for these differences were large. Participants also looked more frequently beyond the previous clause in the sentence than stayed within the clause outside the preceding word/phrase ($Z = 2.26$, $p = .02$). The size of this difference, however, was found to be small ($r = .41$).

The results for below-clause revisions followed similar patterns to what was observed for revisions below the word and at the word level. Participants' eye fixations remained within the previous word/phrase with the greatest frequency, followed by visits to parts of the preceding paragraph, sentence, and clause. Participants looked off-screen as often as they viewed the previous word/phrase, and their eye gazes remained at the point of inscription only a small number of times. A Friedman test yielded a significant overall effect for location of eye movements at textual units, $\chi^2$ (3, $N = 30$) = 31.97, $p < .01$. Follow-up Wilcoxon Signed Rank tests found that, when revisions involved smaller units than a clause, participants' eye gazes remained significantly more frequently within the word/phrase than the previous clause ($Z = 4.63$, $p < .01$, $r = .85$), sentence ($Z = 3.73$, $p < .01$, $r = .68$), and paragraph ($Z = 2.93$, $p < .01$, $r = .53$). In addition, the tests indicated that eye fixations were more frequent outside the sentence in the paragraph than in the sentence beyond the preceding clause ($Z = 2.49$, $p = .01$, $r = .45$). The effect sizes were in the medium to large range.

Substantially fewer revisions were made at the clause level and above than lower textual units. Less than half of the participants viewed any of the interest areas before revising a clause or a longer unit. The Friedman test, which was conducted to test whether there were differences in the location of eye movements before participants

revised at the clause level or above, yielded no significant overall effect for location of eye gazes at textual units, $\chi^2$ (3, $N = 24$) = 6.45, $p = .09$.

Finally, on the few occasions when participants revised a whole sentence or larger textual unit, they most often visited parts of the text that were outside the previous sentence they had composed. A Friedman test confirmed that there was an overall effect for location of eye fixations at textual units, $\chi^2$ (3, $N = 18$) = 20.11, $p < .01$. According to Wilcoxon Signed Ranks tests, when participants revised at the sentence level or above, they significantly more often viewed areas in the preceding sentence beyond the previous clause than text in the previous clause outside the previous word/phrase ($Z = 2.53$, $p = .01$, $r = .60$), and more frequently visited parts of the preceding paragraph further than the previous sentence than areas within the preceding word/phrase ($Z = 2.35$, $p = .02$, $r = .55$) or clause ($Z = 2.97$, $p < .01$, $r = .70$). The effect sizes ranged from medium to large.

Table 7 provides a summary of the significant patterns for eye-gaze behaviors before revisions.

### STIMULATED RECALL COMMENTS ASSOCIATED WITH REVISIONS AT DIFFERENT LEVELS

Table 8 summarizes the stimulated recall comments elicited to describe participants' thoughts during revision. Contrary to what was found for pausing, participants referred to translation mechanisms more frequently (70%) than to planning processes (14%) in total. While the same pattern was observed for all levels of revision, the proportion of translation-related comments gradually decreased as the level of revision increased. The differences between the percentage of comments on translation and planning were 26%, 18%, 7%, and 2%, respectively, at the single word, below clause, clause and above, and sentence and above levels. In other words, participants referred to translation processes proportionately more frequently when they revised lower than higher textual units.

Moving on to the distribution of subprocesses, overall, the majority of planning comments concerned planning content (88%), and most of the translation comments referred to lexical encoding (52%). For planning, similar patterns were observed across revision levels. However, the distribution of translation-related comments was found to

TABLE 7.    Summary of significant patterns for eye-gaze behaviors before revisions

| | |
|---|---|
| Below word | word/phrase > clause, sentence, paragraph; paragraph > clause |
| Word | word/phrase > clause, sentence, paragraph; clause > sentence |
| Below clause | word/phrase > clause, sentence, paragraph; paragraph > sentence |
| Clause and above | n/a |
| Sentence and above | sentence > clause, paragraph > word/phrase |

*Note:* > indicates a significantly larger number of views at a certain textual unit.

TABLE 8.   Reasons for revision: Number of stimulated recall comments by level of revision

| Pause location | Planning | | | Translation | | | | No recall | Total[b] |
|---|---|---|---|---|---|---|---|---|---|
| | Content | Organization | All[a] | Lexical retrieval | Syntactic encoding | Cohesion | All[a] | | |
| Below word | 1 | 0 | 1 (0%) | 13 | 4 | 2 | 21 (7%) | 2 (1%) | 24 (8%) |
| Single word[c] | 5 | 2 | 7 (2%) | 62 | 27 | 12 | 87 (28%) | 16 (5%) | 110 (35%) |
| Below clause | 17 | 0 | 17 (5%) | 32 | 23 | 12 | 71 (23%) | 22 (7%) | 110 (35%) |
| Clause and above | 11 | 0 | 11 (3%) | 8 | 16 | 2 | 30 (10%) | 6 (2%) | 47 (15%) |
| Sentence and above | 4 | 3 | 7 (2%) | 0 | 4 | 3 | 12 (4%) | 5 (2%) | 24 (8%) |
| Total | 38 | 5 | 43 (14%) | 115 | 74 | 31 | 221 (70%) | 51 (16%) | 315 (100%) |

[a]Values for subcategories do not necessarily add up to the total, given that some comments were not specific enough to allow for further subcategorization.
[b]Due to rounding some totals do not add up to 100.
[c]One full word added, deleted, or substituted.

vary according to the level of revision: the percentage of comments on syntactic coding, as compared to lexical retrieval, grew as textual units increased (below word: 21%, single word: 27%, below clause: 34%, clause and above: 62%, sentence and above: 57%).

In summary, according to the stimulated recall comments, revisions were more often concerned with translation than planning-related processes at all levels of revision, but participants referred to translation-related process with proportionately lower frequency when they revised higher textual units such as clauses and sentences.

## DISCUSSION

### PAUSING BEHAVIORS AND UNDERLYING COGNITIVE PROCESSES

Our first research question asked what cognitive writing processes underlay pauses at different textual locations, as reflected in the eye-gaze behaviors and stimulated recall comments of L2 writers. The eye-tracking data revealed that, when participants paused between words, their eye gazes were most likely to visit areas outside the word/phrase preceding the point of inscription but stay within the previous clause. In parallel, during between-sentence pauses, participants were most probable to look beyond the clause but not further than the sentence before the inscription point. According to the stimulated recall comments, participants tended to be more concerned with translation- than planning-related processes when they paused within and between words. In contrast, they recalled focusing more on planning as compared to translation during pauses between sentences. Additionally, Révész, Michel et al. (2017), using the same dataset, found that pause durations increased with increasing textual units, participants pausing longest between sentences followed by pauses between and within words. Taken together, these results indicate that pausing between sentences was more likely to be associated with the rereading of longer stretches of text and engagement in higher-order writing processes such as planning content, whereas pauses between words tended to involve looking back at shorter textual units and engaging in lower-order writing processes including lexical retrieval and syntactic encoding.

These findings are well aligned with the results of Révész, Kourtali et al. (2017) and those of Chukharev-Hudilainen et al. (2019). Révész, Kourtali et al. (2017) also concluded, employing keystroke logging and stimulated recall, that pauses occurring before the production of longer textual units were more likely to reflect higher-level writing processes. In Chukharev-Hudilainen et al.'s study, participants were likewise found to look back in their texts when they paused between larger textual units. Importantly, however, through the triangulation of keystroke logging, eye-tracking, and stimulated recall data, we provided evidence for these patterns based on a single dataset in the current study, allowing for drawing more valid inferences about the processes underlying pausing behaviors.

A finding contrary to our expectations was that, for within-word pauses, no difference emerged in the frequency with which participants viewed various textual units. One explanation for this may be that, because of the relatively high pause threshold of 2 s adopted in the study (cf., Van Waes & Leijten, 2015), we did not capture some of the lower-level writing processes that participants carried out (e.g., retrieving spelling, morphosyntactic encoding). Probably, these shorter pauses, potentially involving below-

word level typographical and linguistic encoding processes, would have been associated with more local eye movements closer to the point of inscription. Another possible account may be related to our observation during data collection that a considerable number of writers engaged in hunt-and-peck writing. Hunt-and-peck writers mostly view the keyboard while composing, and often produce considerably large chunks of text before rereading what they have written (Leijten & Van Waes, 2013). Thus, this type of writers, unlike monitor gazers who primarily look at the screen while they write, might have been less likely to look at the screen during pauses within lower textual units.

## REVISION BEHAVIORS AND UNDERLYING COGNITIVE PROCESSES

Our second research question was concerned with exploring the cognitive processes underlying different levels of revision, that is, whether the revision involved a change below the word level, at the word level, below the clause level, at the clause level or above, or at the sentence level and above. The analysis of the eye-gaze behaviors indicated that participants' eye gazes were most likely to remain within the previous word/phrase before they revised lower textual units (lower than a word, a word, and lower than a clause). However, prior to revising an entire sentence or a longer stretch of text, they were most probable to look at areas beyond the clause in the sentence or further than the sentence at the inscription point. It is also noteworthy that participants were considerably more likely to look off-screen preceding lower- than higher-level revisions. The stimulated recall comments uncovered that participants were more frequently concerned with translation- than planning-related processes regardless of level of revision. However, the proportion of comments on planning, as compared to translation, increased as larger textual units were revised. Overall, these results show that, when participants made lower-level revisions, they predominantly focused on linguistic issues, and, prior to making a lower-level revision, their eyes tended to remain off-screen or fixate within the textual unit they were about to revise. Higher-level revisions, although more often concerned with language problems as well, were more probable to focus on planning-related issues than lower-level revisions, and, before a higher-level revision, participants' eye gazes were most likely to remain on-screen and fixate on the area to be revised.

These results are largely consistent with those of previous L2 research on revision behaviors. Révész, Kourtali et al. (2017) also observed that, while most of their participants' stimulated recall comments focused on translation across all levels of revision, an increasing proportion of planning-related comments occurred as larger textual units were revised. Keystroke logging studies of L2 writing, in general, show that L2 writers make more language- than content-focused revisions (e.g., Barkaoui, 2016; Stevenson et al., 2006). However, the extent of the difference in the distribution of content revisions versus language revisions seems to vary across studies. In the present experiment, the stimulated recall participants recalled focusing on linguistic issues approximately five times more frequently than on content. A similar distribution of content- versus language-oriented revisions was observed in Stevenson et al. (2006), but Barkaoui (2016) found that participants overall made only about three times as many language- as content-focused precontextual changes. This discrepancy in findings might be related to a difference in the amount of online planning that participants had available when composing their essays, with less online planning leading to a decrease in focus on

linguistic encoding (Ellis & Yuan, 2004). In our study, participants were given 40 min to complete the writing task, and the expected word count was 250 words. The time limit was 30 min in Barkaoui's and Stevenson et al.'s research, but the former required participants to produce a 300-word essay, whereas the latter had no set word count. The greater time pressure in Barkaoui's experiment probably left writers with fewer attentional resources to allocate to translation processes.

An intriguing finding emerging from our data concerns the difference in off-screen views preceding lower- and higher-level revisions. One way to account for the considerably higher percentage of off-screen eye gazes before lower-level revisions is to consider the influence of hunt-and-peck writing (Leijten & Van Waes, 2013). Hunt-and-peck writers might have been able to revise lower-level textual units without rereading them on the screen, as rehearsing shorter textual units is less taxing for working memory. In contrast, maintaining larger chunks of text active in working memory is more demanding due to capacity limitations. Therefore, when monitoring their evolving text, hunt-and-peck writers probably had to reread longer textual units before making the decision to revise.

## LIMITATIONS AND FUTURE RESEARCH

In discussing the results of the study, it is also necessary to recognize the limitations of the research. One limitation concerns the relatively long pause threshold (2 s) we adopted. Although researchers have traditionally employed a pause threshold of 2 s in L2 writing and, hence, the use of this threshold aids the comparability of our research to previous L2 studies, adding a shorter threshold would have better enabled us to capture lower-level writing processes (e.g., Baaijen et al., 2012; Van Waes & Leijten, 2015). There are also inherent limitations associated with the use of the stimulated recall methodology (Gass & Mackey, 2017). Owing to memory loss, for example, it is unlikely that participants were able to recall all the thoughts they had while writing. The study would also have profited from the use of a higher-precision eye-tracker, which would have allowed for a more accurate evaluation of eye-gaze behaviors. Future research on L2 writing behaviors could also use technology that tracks keystroke logging and eye-gaze data simultaneously (e.g., Chukharev-Hudilainen et al., 2019). This would potentially permit researchers to obtain a wider range of quantitative measures describing eye movements during pauses and before revisions. In future studies of L2 writing, it would also be interesting to explore relationships between pausing and revision behaviors, given that these two phenomena often co-occur during the writing process (Baaijen et al., 2012). Additional fruitful venues for further research would be to investigate whether the patterns found here apply to other proficiency levels, task types, and L1 and L2 groups, as our research was restricted to advanced L2 writers, a single argumentative essay, and Mandarin users of L2 English. If the results obtained here were to be confirmed in future studies, they could be used as a basis for diagnosing areas of writing difficulty. For example, depending on the distribution of pause locations and levels of revisions (e.g., extensive pausing and revisions at lower textual units), L2 instructors could tailor instruction to meet students' needs (e.g., greater focus on linguistic encoding in writing classes).

Future research would also benefit from applying the combination of the techniques utilized here to address further questions in writing research. The joint use of keystroke logging, eye-tracking, and verbal protocols would appear particularly helpful to examine the processes involved in source-based writing, where writers are required to incorporate content from sources such as images and/or written or oral texts (see Leijten, Van Waes, Schrijver, Bernolet, & Vangehughten, 2019). For example, the eye-gaze recordings would enable researchers to gather direct evidence about how much time writers spend viewing the source(s), and how often they switch between the source(s) and their evolving text. This information, together with keystroke logs and comments from verbal protocols, would assist in tapping source-based writing processes more thoroughly.

**CONCLUSION**

The purpose of the current study was to examine the cognitive processes underlying L2 pausing and revision behaviors during L2 writing. Specifically, our aim was to shed light on the cognitive processes associated with pauses at various textual locations and different levels of revision. The methodological innovation of our study was to employ stimulated recall, keystroke logging, and eye-tracking methodologies in combination to examine different types of pausing and revision phenomena. We found that, when participants paused between sentences, they were more likely to look back on longer texts and engage in higher-order writing processes. In contrast, during pauses within and between words, they tended to view areas closer to the inscription point and be involved in lower-order writing processes. Before making a revision, participants most frequently visited the area that they later revised or, in the case of lower-level revisions, remained off-screen. Revisions, in general, were more probable to focus on language- than content-related issues, but the difference in the proportion of comments on language and content decreased as the level of the revised textual unit increased. These results are well aligned with patterns emerging from previous research. However, through triangulating stimulated recall, keystroke logging, and eye-tracking data, we were able to confirm these patterns based on a single dataset, affording more valid conclusions about the processes underlying pausing and revision behaviors. In general, the study confirmed that the application of these three data sources together allows for obtaining a more complete picture of the writing process than the use of a single technique would make possible.

**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 M. Torrance, L. Van Waes, & D. Galbraith (Eds.), *Writing and cognition: Research and applications* (pp. 55–65). Amsterdam, The Netherlands: Elsevier.

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

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.

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

Broekkamp, H., & van den Bergh, H. (1996). Attention strategies in revising a foreign language text. In G. Rijlaarsdam, H. van den Bergh, & M. Couzijn (Eds.), *Theories, models and methodology in writing research* (pp. 170–181). Amsterdam, The Netherlands: Amsterdam University Press.

Chanquoy, L., Foulin, J.-N., & Fayol, M. (1990). Temporal management of short text writing by children and adults. *Cahiers de Psychologie Cognitive/Current Psychology of Cognition*, *10*, 513–540.

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.

Damian, M. F., & Stadthagen-Gonzalez, H. (2009). Advance planning of form properties in the written production of single and multiple words. *Language and Cognitive Processes*, *24*, 555–579.

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

Elola, I., & Mikulski, A. (2013). Revisions in real time: Spanish heritage language learners' writing processes in English and Spanish. *Foreign Language Annals*, *46*, 646–660.

Faigley, L., & Witte, S. (1981). Analyzing revision. *College Composition and Communication*, *32*, 400–414.

Flower, L., & Hayes, J. R. (1980). The cognition of discovery: Defining a rhetorical problem. *College Composition and Communication*, *31*, 21–32.

Galbraith, D. (2009). Cognitive models of writing. *German as a Foreign Language*, *2009*(2–3), 7–22.

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*(2), 469–506. doi:10.1111/lang.12280.

Gass, S. M., & Mackey, A. (2017). *Stimulated recall methodology in second language research*. New York, NY: Routledge.

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). Mahwah, NJ: Erlbaum.

Hayes, J. R. (2012). Modeling and remodeling writing. *Written communication*, *29*(3), 369–388.

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). Mahwah, NJ: Lawrence Erlbaum Associates.

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

Kormos, J. (2012). The role of individual differences in L2 writing. *Journal of Second Language Writing*, *21*, 390–403.

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

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

Li, J. (2006). The mediation of technology in ESL writing and its implications for writing assessment. *Assessing Writing*, *11*, 5–21.

Lindgren, E., & Sullivan, K. P. H. (2006a). Analysing online revision. In K. P. H. Sullivan & E. Lindgren (Eds.), *Computer key-stroke logging: Methods and applications* (pp. 157–188). Oxford, UK: Elsevier.

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

Manchón, R., Roca de Larios, J., & Murphy, L. (2007). A review of writing strategies: Focus on conceptualizations and impact of first language. In A. Cohen & E. Macaro (Eds.), *Language learner strategies* (pp. 229–250). Oxford, UK: Oxford University Press.

Matsuhashi, A. (1981). Pausing and planning: The tempo of written discourse production. *Research in the Teaching of English*, *15*(2), 113–134.

Matsuhashi, A. (1987). Revising the plan and altering the text. In A. Matsuhashi (Ed.), *Writing in real time: Modeling production processes* (pp. 197–223). Norwood, NJ: Ablex.

Medimorec, S., & Risko, E. F. (2017). Pauses in written composition: On the importance of where writers pause. *Reading and Writing*, *30*, 1267–1285.

Plonsky, L., & Oswald, F. L. (2014). How big is "big"? Interpreting effect sizes in L2 research. *Language Learning*, *64*, 878–912.

Porte, G. K. (1996). When writing fails: How academic context and past learning experiences shape revision. *System*, *24*, 107–116.

Porte, G. K. (1997). The etiology of poor second language writing: The influence of perceived teacher preferences on second language revision strategies. *Journal of Second Language Writing*, *6*, 61–78.

Révész, A., Kourtali, N., & Mazgutova, D. (2017). Effects of task complexity on L2 writing behaviors and linguistic complexity. *Language Learning*, *67*, 208–241.

Révész, A, Michel, M., & Lee, M. (2017). Investigating IELTS Academic Writing Task 2: Relationships between cognitive writing processes, text quality, and working memory. *IELTS Research Reports Online Series*, 2017/3.

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.

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

Scardamalia, M., & Bereiter, C. (1987). Knowledge telling and knowledge transforming in written composition. *Advances in Applied Psycholinguistics*, *2*, 142–175.

Schilperoord, J. (1996). *It's about time: Temporal aspects of cognitive processes in text production.* Amsterdam, The Netherlands: Rodopi.

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

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.

van Hell, J. G., Verhoeven, L., & van Beijsterveldt, L. M. (2008). Pause time patterns in writing narrative and expository texts by children and adults. *Discourse Processes: A Multidisciplinary Journal*, *45*, 406–427.

Van Waes, L., & Leijten, M. (2015). Fluency in writing: A multidimensional perspective on writing fluency applied to L1 and L2. *Computers and Composition: An International Journal*, *38*, 79–95.

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*, 829–853.

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). Amsterdam, The Netherlands: Elsevier.

Whalen, K., & Ménard, N. (1995). L1 and L2 writers' strategic and linguistic knowledge: A model of multiple-level discourse processing. *Language Learning*, *45*, 381–418.