

The temporal distribution of cognitive writing processes and its interaction with task complexity during computer-assisted collaborative writing

Language Teaching Research

1–28

© The Author(s) 2025



Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/13621688251352287

journals.sagepub.com/home/ltr**Xin Rong** 

University College London, UK

Andrea Révész

University College London, UK

Abstract

This study examined how timing influences writing behaviors and associated cognitive activities in second language users during computer-assisted collaborative writing and whether task complexity mediates this relationship. The study involved 56 Chinese participants with English proficiency levels at CEFR B2 and C1. They were randomly arranged into 28 pairs. Each pair completed two counterbalanced writing tasks in a reading-to-write format, differing in cognitive complexity. The simple task version involved summarizing a single text, whereas the complex task version required writing a summary of three texts. Keystroke logging software tracked the participants' typing behaviors during the tasks. In addition, eight pairs were selected through stratified sampling for stimulated recall interviews immediately after completing the second task. Analyses using linear mixed-effects models revealed significant time effects on the duration and frequency of within-word pauses and revisions at various levels, as well as two interaction effects between time and task complexity for between-sentence pause length and between-sentence pause frequency. These results, together with stimulated recall comments, highlight the dynamic interplay between time- and task-related factors during the collaborative writing process.

Keywords

Collaborative writing, writing behaviors, cognitive activities, temporal dimension, keystroke logging, stimulated recall

Corresponding author:

Xin Rong, University College London, Gower Street, London WC1E 6BT, UK.

Email: xin.rong.19@alumni.ucl.ac.uk

I Introduction

Collaborative writing refers to the activity of learners working together to produce a text throughout the writing process (Storch, 2013). This writing practice has accumulated increasing attention over the past decades as a tool to promote learners' writing development. As compared with individual writing, collaborative writing has been found to promote the accuracy of text production (Elabdali, 2021). In addition, it creates language learning opportunities that are thought to benefit second language (L2) learning processes (Storch, 2021). Nevertheless, little is known about the cognitive processes involved in collaborative writing, as most of previous process-oriented research focused on the social aspects of collaboration (Storch, 2019). However, since collaborative writing involves both social and cognitive elements, it is also important to examine learners' writing behaviors and the underlying cognitive activities. A better understanding of these processes can help identify writing challenges and support the development of more personalized L2 instruction (Révész et al., 2019). Therefore, the present study adopted a cognitive perspective to investigate L2 users' writing activities in collaborative writing.

In this study, collaborative writing was conducted among university students, given its growing relevance to this population. Collaborative writing tasks are increasingly integrated into university programs, requiring students to produce joint pieces of writing, reflecting real-life writing tasks such as coauthoring project reports and articles. Collaborative writing can take place either face-to-face or via computer-mediated communication. We chose to focus on face-to-face computer-assisted collaborative writing because it has been shown to promote more effective collaboration compared with computer-mediated settings (e.g., Rouhshad et al., 2016). Our research specifically aimed to explore how cognitive processes differ across writing periods and how task complexity may influence the temporal distribution of collaborative writing processes. We focused on task complexity, the inherent cognitive demands of tasks (Robinson, 2001), with a view to providing language teachers with insights on how to sequence collaborative writing tasks effectively, considering the potentially different cognitive load imposed by various versions of collaborative writing tasks. We chose to concentrate on integrated writing (i.e., creating a text using source materials) rather than independent writing (i.e., solely relying on one's own resources) to better reflect real-world academic writing practices. To gain a fine-grained picture of writing processes, we adopted a mixed-methods approach, integrating keystroke logging and stimulated recall data, being among the first to employ these techniques in combination to investigate collaborative writing processes.

II Background

I Theoretical background

Among cognitive models of writing (e.g., Hayes, 2012), the current study draws upon Kellogg's (1996) and Rijlaarsdam and van den Bergh's (1996) theoretical work on writing processes. Kellogg's writing model, though initially developed to describe first language (L1) writing, is well-suited for research on L2 writing processes (Révész & Michel, 2019). Unlike other writing models, it places greater emphasis on linguistic

encoding processes, which are a notable challenge in L2 writing (Révész et al., 2019). The model identifies three systems of text production: formulation, execution, and monitoring. Formulation involves planning structure and content, retrieving information, and translating ideas into language. Execution controls the motor skills needed to write or type. In monitoring, writers review and edit their text to correct any discrepancies or errors. The model depicts writing as an interactive and recursive process, with the three systems working together and, thus, making it possible for each writing activity to occur at any time. However, the writing process is more prone to interruptions in L2 than L1 writing, as breakdowns in the translation process are more likely to occur, due to L2 learners' limited processing capacity (Leow, 2015) and L2 knowledge.

One limitation of Kellogg's model is its lack of detail regarding how these cognitive processes occur in parallel throughout the writing process. Rijlaarsdam and van den Bergh (1996) suggest that the likelihood of engaging in a particular cognitive writing activity change at different stages of writing. In other words, specific behaviors are more likely to happen at certain times rather than others. For instance, planning activities are more common in the early stages of writing. In addition, the same cognitive activity may play different roles at various points in the writing process. The function of an activity at any given time depends on the context and the preceding activities that have triggered it (Rijlaarsdam & van den Bergh, 2006).

2 Writing processes and the time course of writing

Inspired by Rijlaarsdam and van den Bergh's (1996) work, a growing body of research has examined the temporal nature of L2 writing processes. Earlier studies investigated the evolution of writers' cognitive activities over time, primarily through participants' concurrent verbal reports (Manchón & Roca de Larios, 2007; Roca de Larios et al., 2001, 2008; Tillema, 2012; Van Weijen, 2009). L2 writers were found to engage in planning-related activities predominantly at the beginning of a writing task and focus on formulation during the middle stages. However, findings on the temporal distribution of revision activities were less consistent. Some researchers (e.g., Roca de Larios et al., 2008) observed a gradual increase in revision behaviors over time, whereas others (e.g., Tillema, 2012) reported that revision activities were evenly distributed throughout the writing process.

More recent studies have utilized keystroke logging software, either alone or in conjunction with other methods, to analyze learners' writing behaviors. Van Waes and Leijten (2015) found that L2 writers' text production, measured by the number of characters typed, progressively decreased over time. In a study by Xu and Qi (2017), university students with higher writing abilities produced longer and less-frequent pauses during the initial writing phase, whereas less skilled writers exhibited longer and less-frequent pauses in the second interval. These results were partially supported by Barkaoui (2019), who observed that the initial writing period involved less-frequent but longer pauses, regardless of task type or L2 proficiency. In addition, Michel et al. (2020) noted that the fewest characters were produced in the final writing period, with a higher frequency of pauses occurring in the middle periods.

The findings on revision behaviors were less consistent. In Barkaoui's (2016) study, students produced the most revisions during the middle period of writing. When these

revisions were further categorized by their location, slightly different conclusions emerged. Precontextual revisions (i.e., revisions made at the point of inscription) were found to occur more often in the middle period, whereas contextual revisions (i.e., revisions made prior to the point of inscription) were more frequent in the final writing period. Conversely, Gánem-Gutiérrez and Gilmore (2018) observed a relatively stable trend in revising behavior throughout the writing process. In a study of L2 Chinese learners' revision behaviors, Lu and Révész (2021) found that overall and precontextual revisions were more common in the middle periods, whereas the frequency of contextual revisions increased from the beginning to the end of the writing process. Similarly, Révész et al. (2023) discovered that participants tended to make more revisions in the later stages of writing.

Collectively, these studies provide empirical support for the dynamic nature of the L2 writing process. In the initial writing stages, writers were found to focus on planning the organization and content of their texts, as indicated by their verbal comments, and longer but less-frequent pauses recorded by keystroke logging. Verbal protocol data suggest that formulation typically occurs in the middle stages, which aligns with the observation of shorter, more-frequent pauses and local revisions occurring at the point of writing. In most studies, the final periods of writing saw an increase in revisions made to previously written content, suggesting increasing monitoring activities as the writing process progressed.

3 Task complexity, collaborative writing, and the time course of writing

Although previous studies have provided empirical insights into the dynamic processes of L2 writing, they have generally overlooked how manipulations in task design might influence the writing process. In contrast, task effects have received significant theoretical attention in L2 speaking research, particularly concerning the concept of task complexity (Robinson, 2001).

Inspired by this concept, a few studies have begun to investigate the effect of task complexity on L2 writing processes. In Révész et al.'s (2017) study, task complexity was operationalized as whether extra ideas were available in an argumentative essay. Participants in the simple task exhibited less-frequent pauses but more-frequent revisions and reported fewer cognitive activities related to planning. Similarly, Jung (2020) demonstrated that, when content support was provided, participants paused less frequently and for shorter durations and produced fewer inserts. However, neither study explored how task complexity moderated these differences across different periods of the writing process.

The temporal dimensions of learners' writing processes in collaborative writing have received limited attention. In Mak and Coniam's (2008) exploratory study, a group of four students' writing change functions were recorded across three phases. However, due to asynchronous collaboration on Wiki, each phase lasted two weeks. Other studies divided the writing process into three phases including planning, composing, and revising and reported the length and/or proportion of each phase (McDonough et al., 2016; Teng & Huang, 2021; Wigglesworth & Storch, 2009). These studies also examined the occurrence of writing activities by coding and categorizing pair dialogues. Nonetheless,

it is noteworthy that the division still treats writing as a linear process, without accounting for the possibility that the same writing activity might occur at different times throughout the writing process (Rijlaarsdam & van den Bergh, 1996). In addition, these coded writing activities were not analyzed in relation to the specific time at which they took place.

Little is known about the effect of task complexity on collaborative writing processes. One exception is the study conducted by Hsu (2020), which explored the influence of task complexity on pair dynamics. In the study, dyads performed two asynchronous writing tasks that varied in their level of reasoning demands. In contrast to the researcher's expectations, the findings did not indicate a significant effect of task complexity on pairs' interaction process. Although this study offered valuable insights, it did not address how task complexity might shape the temporal distribution of writing behaviors and cognitive activities involved in collaborative writing.

III Research questions

Inspired by previous theoretical and empirical work, this study explored the effect of writing period and its interaction with task complexity on writing behaviors and underlying cognitive activities in the context of computer-assisted collaborative writing. We proposed two research questions.

In computer-assisted collaborative writing:

1. To what extent does writing period influence writing behaviors and cognitive activities?
2. To what extent does task complexity mediate the influence of writing period on writing behaviors and cognitive activities?

We analyzed writing behaviors in terms of speed fluency, pausing, and revision. Cognitive activities were reported in stimulated recall interviews. To consider the influence of writing period, we split each writing session into five equal time periods. Task complexity was operationalized as the need to summarize one complete text (the simple task version) or three texts (the complex task version).

IV Method

I Design

The dataset for this study is part of a larger research project (Rong & Révész, 2025) that includes both collaborative and individual writing samples. In this study, we focus solely on the collaborative writing data of 56 participants organized into 28 randomly assigned pairs. Each pair completed two collaborative writing tasks on two different topics: nuclear power and driverless cars. We designed a simple and complex version of a reading-to-write task and presented the resulting task versions with task complexity and topic counterbalanced across pairs. To capture participants' writing behaviors, we used the keystroke logging tool Inputlog (Version 8.0.0.17, Leijten & Van Waes, 2013), which

recorded participants' keyboard and mouse activities. In addition, we employed screen recording software to capture the laptop screen. Data of participants' cognitive activities associated with the writing behaviors were collected through stimulated recall interviews. Eight pairs, selected using a stratified sampling approach taking task complexity and topic into account, participated in the interview immediately after their second task performance.

2 Participants

Given our G*Power calculations (Faul et al., 2007) and within-participants design, we aimed for a sample size of 56 participants to ensure adequate statistical power and to equally cover all counterbalanced conditions. Initially, we recruited 89 Chinese learners of L2 English but excluded 28 participants based on their proficiency and typing test results. The remaining 61 participants were eligible for collaborative writing tasks. To meet the requirements of the counterbalanced design, five additional participants were randomly excluded from the study. Our final sample comprised 53 women and 3 men, with ages ranging from 19 to 33 years ($M = 24.14$, $SD = 3.10$). All participants were university students at a London-based university in the United Kingdom. Most were postgraduate students, with one undergraduate and three doctoral students. Over half of the participants ($n = 35$) were enrolled in programs of teaching English to speakers of other languages and applied linguistics.

3 Instruments

a Proficiency test. We assessed participants' English proficiency using a practice version of the Cambridge English Advanced (CAE) test. Chinese participants were less familiar with this test, which helped minimize any potential inflation of proficiency due to repeated test-taking or coaching. Only the reading and writing sections of the CAE test were administered, as these skills were deemed most pertinent to integrated writing task performance. The reading section was marked by the first author, whereas participants' performance on the two tasks in the writing section was independently assessed by two raters. The final writing score was calculated as the average of the two raters' scores. Intraclass correlation coefficients (ICCs) indicated good consistency across the raters: Part 1, $ICC(3, 2) = .87$ ($p < .001$) and Part 2, $ICC(3, 2) = .80$ ($p < .001$) (Koo & Li, 2016). We excluded participants whose scores fell outside the B2 to C1 range in either the reading or writing sections on The Cambridge English Scale.

b Typing test. A typing test was used to assess participants' keyboarding skills. Each participant completed two 2-minute online typing tests using their personal laptops, following a 1-minute practice test. The tests required participants to type passages as quickly and accurately as possible with the correction function disabled. One participant whose result deviated more than three standard deviations from the group mean was excluded. The average net typing speed, adjusted for accuracy, was 33.41 words per minute ($SD = 7.75$) for the final sample.

c Collaborative writing tasks. As mentioned previously, collaborative writing tasks were designed within integrated writing. Considering the target participants being mostly postgraduates, integrated writing would be more frequent in the academic environment than independent writing, therefore leading to increased ecological validity (i.e., the generalizability of findings obtained in research settings to the real world; Orne, 1962). From a practical perspective, the target participants were expected to be familiar with integrated writing and to possess a basic understanding of how to perform an integrated writing task. Our expectation was based on the fact that most participants were postgraduates and the data collection took place during the second half of the first academic term.

Task complexity was operationalized as the simple and complex versions of a reading-to-write task. The task version designed to be less complex required participants to summarize a single article. In contrast, the task designed to be more complex asked participants to write a summary by integrating information from three texts. The two source texts for the simple task version focused on different topics: nuclear power and driverless cars. However, both texts were similar in terms of genre, text structure, and linguistic complexity. Both were expository articles, and each comprised nine paragraphs, including an introductory paragraph with a thesis statement of the gist of the text, seven paragraphs discussing four advantages and three disadvantages related to the topic, and a concluding paragraph for the current development of the topic. The two texts were then linguistically modified. The results (see Table S1 in the online supplemental material for the text statistics) indicated that the two texts achieved comparable lexical, syntactic, and discourse complexity, as well as readability and word count.

To enhance the cognitive demands of the complex task version while maintaining comparable complexity and readability of the source texts used in both task versions, we reorganized the paragraphs of each text used in the simple task version into three shorter texts with minor modifications. Specifically, we removed the thesis statement and divided the seven body paragraphs into three texts, each addressing one disadvantage and one or two advantages of the topic. We also made adjustments to the introductory and concluding paragraphs to balance the number of paragraphs and word counts across the three text segments. All text versions are available in S2 in the online supplemental material.

The pairs who met the selection criteria of the proficiency test and the typing test carried out the collaborative writing tasks in a research lab. Pairs underwent pretask modeling before the first collaborative writing task. As most participants ($n = 46$) had not had the experience of collaborative writing prior to the experiment, pretask modeling was necessary to help them become acquainted with collaborative writing. It was conducted in the participants' L1 as a four-step activity adapted from Chen and Hapgood (2021). The process began with a communicative activity in which participants discussed London tourist attractions they had both visited. In the second step, participants first shared their understanding of collaborative writing based on their language learning and, if applicable, teaching experiences. Following this, Storch's (2019) definition of collaborative writing was presented. In the third step, Storch's (2016) four interaction patterns were introduced to emphasize the need for true collaboration, where each participant has an equal role in the decision-making process (i.e., equality) and the engagement in each

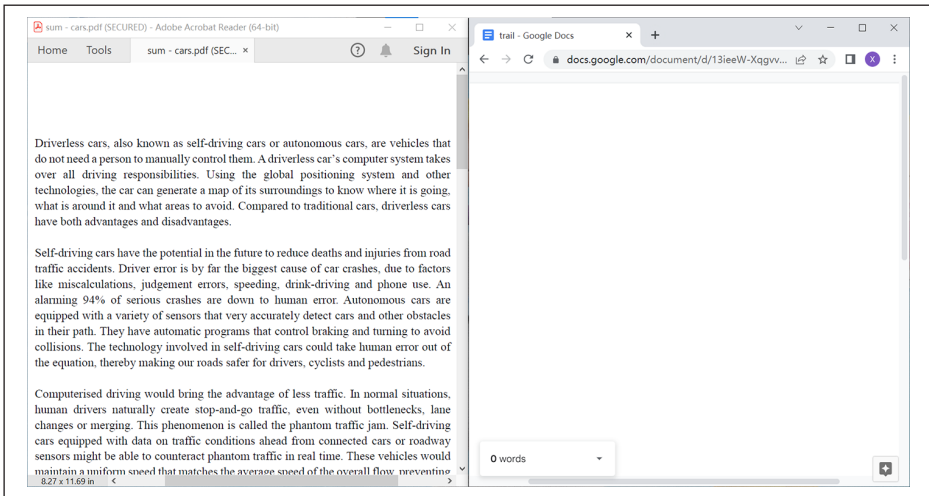


Figure 1. Screenshot of the task setup.

other's contribution (i.e., mutuality). Finally, participants practiced a brief collaborative writing task in 10 minutes, which involved creating a 50-word recommendation for a tourist attraction in London they had discussed during the icebreaker activity to their friends.

During collaborative writing tasks, participants sat face-to-face with each other and worked on individual laptops provided by the researchers. One side of the laptop screen displayed the source material(s), whereas the other side presented a Google Docs page. This setup, as shown in Figure 1, enabled participants to collaboratively take notes, develop outlines, and compose their texts in real time. The writing tasks were framed as a joint effort akin to magazine editors collaborating to produce an article for their readers. Participants were instructed to draft content intended for a general readership without adding personal perspectives. They were also told not to copy sentences from the source texts.

To determine the word count and time limit for the tasks, a pilot study was conducted with 14 participants similar to the target group. Each participant completed a simple task version individually, resulting in a median text length of 235 words and a median completion time of 49 minutes. Given that collaborative tasks typically take 1.5 times longer than individual tasks (Wigglesworth & Storch, 2009), participants in pairs were instructed to produce between 225 to 275 words within 75 minutes. Pairs were allowed to use L1 for peer interaction during collaborative writing tasks. Nonetheless, participants could switch between languages at any time if they wished to do so.

d Questionnaires. We administered a background questionnaire to gather demographic data. After each task performance, participants also responded to a post-task questionnaire (see S3 in the online supplementary file for the questionnaire) with 9-point Likert scale items (Révész, 2014). The questionnaire first gauged participants' familiarity with the topics and content covered in the texts. In addition, one item evaluated participants'

perceived mental effort involved in reading the source text(s). Given the controlled complexity, readability, and length of the source text(s), we would expect minimal variation in participants' perceived reading mental effort across task versions. Another item assessed participants' perceived mental effort during writing, as we wanted to examine whether, as intended, they exerted greater mental effort in summarizing three texts than in summarizing one text. Participants' perceived task difficulty was also evaluated.

e Stimulated recall interview. The stimulated recall interviews were conducted by the first author to explore the cognitive activities underlying participants' writing behaviors. To overcome the influence of pairs' potential interaction on the recall data, interviews were conducted individually in a consecutive order with two participants. The researcher presented participants with screen recordings of their writing processes as stimuli and instructed them to verbalize what they were thinking during the writing process. The instructions were adapted from Gass and Mackey (2017) and standardized to ensure the consistency of the procedure. Participants were encouraged to pause the recording at any point they wanted to share their thoughts. In addition, the researcher prompted participants to recall their thoughts when they paused or revised their texts (e.g., "What made you pause/revise at this point?") unless they commented on these behaviors on their own. The interviews were carried out in the participants' L1.

4 Data collection

Figure 2 depicts the data collection schedule. In the first group session, participants completed the ethics procedures, the CAE reading and writing practice test, and the background questionnaire. Shortly after the group session, they took the online typing test. Next, participants who met the selection criteria were randomly assigned to pairs. In the second session, each pair engaged in pretask modeling activities and carried out the first collaborative writing task, followed by the post-task questionnaire. The third session mirrored the structure of the second session, with pairs completing the collaborative writing task on a different topic then completing the post-task questionnaire. The session ended with individual stimulated recall interviews with participants from eight pairs.

5 Data analysis

a Writing behaviors. Before obtaining the writing behavioral data, we split the total time of each writing session into five equal periods (e.g., Michel et al., 2020; Révész et al., 2023). We excluded writing behaviors in the initial planning stage (e.g., taking notes, creating outlines), as they considerably differed from those in subsequent text production stages (Baaijen et al., 2012). However, instances when participants developed their notes and directly incorporated them into the final text were considered part of the text production phase. Because of this, data for the first period were removed from subsequent statistical analyses.

We manually coded participants' writing behaviors in terms of speed fluency, pausing, and revision by using the general analysis produced by Inputlog. Speed fluency was evaluated using two metrics: production rate (the number of characters typed per second) and P-burst length (the number of characters typed between consecutive pauses, Baaijen

Session 1	Session 2	Session 3
Consent form (3 mins)	Pre-task modeling (40 mins)	Collaborative writing task two (75 mins)
Reading test (90 mins)	Collaborative writing task one (75 mins)	Post-task questionnaire (2 mins)
Break (10 mins)	Post-task questionnaire (2 mins)	Stimulated recall interview (avg. 72 mins)
Writing test (90 mins)		
Background questionnaire (4 mins)		
Typing test (8 mins)		

Figure 2. Data collection schedule.

et al., 2012). Pauses were categorized into linear pauses (inactivity during forward progression without cursor movements; Hall et al., 2024) and nonlinear pauses. As each participant was provided with a laptop to type, two Inputlog files were produced in each task. Each file contained nonlinear pauses during which the text was being produced by the other participant on their laptop. We excluded these nonlinear pauses due to their significant variation in duration (several tens of seconds or even minutes). A pause threshold of 2 seconds was applied (Wengelin, 2006), and pause length and frequency were measured within words, between words, between subsentences (with a comma), and between sentences (with a full stop). Frequency was standardized by the number of pauses per 100 keystrokes. Revision behaviors, including deletions, additions, substitutions, and text movements, were analyzed based on the level of linguistic domain involved in the revision: below-word, full-word, below-clause, and clause-and-above (Stevenson et al., 2006). Revisions were also classified as precontextual (revision made at the point of inscription) or contextual (revision made to the already written text) (Lindgren & Sullivan, 2006).

To ensure reliability, a second coder independently coded data from eight sessions (14.3% of the writing sessions), selected through stratified random sampling considering task complexity and topic. Inter-coder reliability was high, with Cohen’s kappa values of .99 for pausing categorization, .95 for level of linguistic domain of revision categorization, and .99 for context of revision categorization.

b Stimulated recall comments. The stimulated recall data were analyzed qualitatively through a four-step process. First, the data were transcribed and divided into segments. Each segment represented a participant’s comment on a specific pausing or revision behavior. In the second step, emergent microcategories were identified. Third, these microcategories were grouped into broader categories guided by Kellogg’s (1996) model and into resource use categories based on Michel et al. (2020). A category related to task requirements was also added to the coding framework (see Table S4 in the online supplemental material for the coding scheme and examples). Finally, the number of

segments assigned to each category and their percentage within each period for each task version were calculated. Pausing and revision-related comments were further broken down by pause location and the linguistic domain and context of revision. Four protocols (25% of the data) were double-coded, resulting in a Cohen's kappa of .80. Disagreements between the coders were addressed through discussion.

6 Statistical analyses

Descriptive statistics were computed in SPSS 29.0.0.0, with outliers identified and trimmed using a threshold of three standard deviations from the group mean. Logarithmic or square root transformation was applied to pausing and revision measures, given that these data were skewed. After normality check, Pearson correlations were conducted for each set of measures tapping a subconstruct of writing behaviors across the two task versions to address potential collinearity and reduce the number of analyses. As no very strong correlations (with correlation coefficients equal to or greater than $\pm .80$; Tolmie et al., 2011) were found between any one measure and every other measure within the same subconstruct in either task condition (see Tables S5–S10 in the online supplemental material for the correlational analyses results), all writing behavior measures were kept in subsequent analyses.

To investigate our research questions, we constructed linear mixed-effects models in RStudio (Version 4.3.2, R Core Team, 2023), utilizing the *lmer* function from the *lme4* package (Version 1.1–35.1, Bates et al., 2015). For all models, the dependent variable was a measure of writing behavior, and the random effects were participant and/or topic. Including topic as a random effect helped control for potential differences in topic familiarity across the source text(s). In models addressing the first research question, writing period served as the fixed effect. For the models investigating the second research question, writing period and its interaction with task complexity were added as extra fixed effects. In each model, participant-by-period or participant-by-task random slopes were also added if model boundary fit could still be retained. To obtain effect sizes, we utilized the *r.squared GLMM* function from the *MuMIn* package (Version 1.47.5, Barton, 2020) to obtain marginal (R^2m) and conditional (R^2c) R squared values, which assessed the variance explained by the fixed effect(s) alone and jointly by the fixed and random effects, respectively. The assumptions of normality and homoscedasticity were confirmed by residual plots using *sjPlot* package (Version 2.8.16, Lüdtke, 2020). The p values for linear mixed-effects models were obtained from the *lmerTest* package (Version 3.1–3, Kuznetsova et al., 2017). The α level was set at .01 to minimize the chance of type I error.

V Results

I Preliminary analyses

Table 1 provides the descriptive statistics for participants' familiarity with the two topics and the contents of the text(s). Table 2 presents the descriptive statistics for participants' perceived familiarity with topics and content of the two task versions, mental effort

Table 1. Descriptive statistics for perceptions of topic and content familiarity by individual participants between topics ($N = 56$).

Rated item	Nuclear power				Driverless cars			
	M	SD	95%CI lower	95%CI upper	M	SD	95%CI lower	95%CI upper
Topic familiarity	5.13	2.24	4.52	5.73	5.71	1.93	5.20	6.23
Content familiarity	5.50	2.05	4.95	6.05	6.07	1.84	5.58	6.56

Table 2. Descriptive statistics for perceptions of topic and content familiarity, reading and writing mental effort, and task difficulty by individual participants across task complexity ($N = 56$).

Rated item	Simple task version				Complex task version			
	M	SD	95%CI Lower	95%CI upper	M	SD	95%CI lower	95%CI upper
Topic familiarity	5.57	2.11	5.01	6.14	5.27	2.10	4.70	5.83
Content familiarity	5.98	1.94	5.46	6.50	5.59	1.97	5.06	6.12
Reading mental effort	4.55	1.75	4.09	5.02	4.89	1.86	4.40	5.39
Writing mental effort	5.13	1.70	4.67	5.58	5.88	1.76	5.40	6.35
Task difficulty	4.38	1.37	4.01	4.74	4.71	1.40	4.34	5.09

required to read the source texts and to write their own texts, and overall difficulty experienced during task performance. According to Table 3, there was no difference in participants’ familiarity with the topics (nuclear power and driverless cars) or the content of the source texts across the two task versions. As expected, neither was the mental effort exerted during reading significantly different between the task versions. However, participants found writing a summary of three texts more mentally demanding than writing a summary of one text, as predicted. Overall task difficulty ratings were also in the anticipated direction, but they did not reach significance. Nevertheless, the mental effort ratings support the validity of our task complexity manipulation.

2 Temporal distribution of writing behaviors and cognitive activities

The first research question explored the extent to which L2 writers show different writing behaviors and cognitive activities across different writing periods. Table S11 in the online supplemental material presents the descriptive statistics of speed fluency, pausing, and revision behaviors by participants across writing periods. Although no statistically significant results were found for speed fluency, writing period significantly predicted some measures of pausing and revision behaviors, as detailed in Table 4 (refer to Table S12 in the online supplemental material for all model results). Turning to the specific results obtained for pausing, we only found significant effects of writing period for within-word pauses. Writers exhibited longer pauses within words in Period 5 as compared with earlier periods and displayed more frequent within word pauses in Period 2 as compared to Period 4. Regarding revision behaviors, writing period showed an impact

Table 3. Results from linear mixed-effects models examining the effects of task complexity on perceptions of topic and content familiarity, reading and writing mental effort, and task difficulty by individual participants ($N = 56$).

Dependent variable	Fixed effect (task complexity)				Random effect			
	Estimate	SE	t	p	R^2m	R^2c	Factor	Variance SD
Topic familiarity	-0.14	0.14	-1.02	.313	< .01	.44	Participant	0.44 0.66
Content familiarity	-0.20	0.14	-1.43	.159	.01	.45	Participant	0.44 0.67
Reading mental effort	0.19	0.14	1.30	.198	< .01	.44	Participant Prompt	0.38 0.62 0.06 0.25
Writing mental effort	0.43	0.14	3.02	.004**	.05	.45	Participant	0.40 0.64
Task difficulty	0.24	0.14	1.69	.097	.01	.42	Participant	0.41 0.64

** $p < .01$.

across nearly all measures, except for clause-and-above revisions. Within-word revisions were more frequent in Period 5 than previous periods. Participants revised full words more often in Period 5 than in Periods 2 and 3 and revised below-clause revisions more frequently in Period 5 as compared with Period 2. The increased revision frequency at various domains also contributed to overall revision being the most frequent in the last period. When it comes to the context where revision behaviors took place, the frequency of precontextual revisions declined between Period 3 and Period 5, whereas contextual revisions increased consistently from Period 2 onward. Although yielding significant results, the effect size for these models was small. Writing period explained between 2% and 7% of the variation in these pausing and revision behaviors.

Next, we investigated the effect of writing period on participants' cognitive activities underlying writing behaviors. Table S13 in the online supplemental material summarizes the number and percentage of stimulated recall comments on pausing-related cognitive activities. Consistent with the trends observed for pausing behaviors, the distribution of pausing-related comments on planning, translation, and resource-use remained relatively stable after the initial period. When examining pause locations, the patterns of these activities also remained consistent throughout the writing process.

Moving onto revision-related cognitive activities, Tables S14 and S15 in the online supplemental material present the number and proportion of stimulated recall comments referring to revision behaviors according to linguistic domain and context, respectively. Similar to pausing, the focus of revision-related stimulated recall comments stayed relatively even across the writing process for both linguistic domain and context. Notably, however, participants made somewhat increased reference to revisions to meet the task requirements during the final period (6%) than previous periods combined (less than 1%).

3 Interaction between writing period and task complexity on writing behaviors and cognitive activities

The second research question examined the extent to which writing period interacted with task complexity to influence writing behaviors and cognitive activities. Table S16 in the online supplemental material presents the descriptive statistics for speed fluency, pausing, and revision behaviors by participants across five writing periods in both the simple and complex task conditions. As indicated in Table 5, we observed significant interaction effects between the writing period and task complexity for two pausing behavior measures: between-sentence pause length between Periods 3 and 5 and between-sentence pause frequency between Period 2 and Periods 3, 4, and 5 (see Figure 3 for illustration, refer to Table S17 in the online supplemental material for all model results). The fixed effects in each model accounted for 7% and 17% of the variance, respectively.

To explore these significant interactions, we separately examined participants' pausing behaviors across writing period for each task version. However, only one model reached significance (Table 6). That is, in the complex task condition, participants exhibited more-frequent between-sentence pauses in Period 2 than Periods 3, 4, and 5, with

Table 4. Significant results from linear mixed-effects models examining the effects of writing period on pausing and revision behaviors by individual participants.

Measure	Reference level	Fixed effect	Estimate	SE	t	p	R ² m	R ² c	Random effect		
									Factor	Variance	SD
Pause length (in seconds, sqrt)											
Within a word (N = 53)											
	Period 2	Period 5	0.62	0.21	2.98	.003**	.07	.07	Prompt	< 0.01	0.03
	Period 3	Period 5	0.60	0.19	3.17	.002**	.07	.07	Prompt	< 0.01	0.03
	Period 4	Period 5	0.72	0.19	3.84	< .001***	.07	.07	Prompt	< 0.01	0.03
Pause frequency (number of pauses per 100 keystrokes, ln)											
Within a word (N = 53)											
	Period 2	Period 4	−0.54	0.16	−3.29	.001**	.04	.33	Participant	0.28	0.53
									Prompt	< 0.01	0.09
Revision frequency (number of revisions per 100 keystrokes, ln)											
Total (N = 56)											
	Period 2	Period 5	0.56	0.12	4.48	< .001***	.05	.24	Participant	0.17	0.41
									Prompt	0.02	0.15
	Period 3	Period 5	0.52	0.12	4.34	< .001***	.05	.24	Participant	0.17	0.41
									Prompt	0.02	0.15
	Period 4	Period 5	0.42	0.12	3.47	< .001***	.05	.24	Participant	0.17	0.41
									Prompt	0.02	0.15
By linguistic domain											
Below word (N = 56)											

By linguistic domain
Below word (N = 56)

(Continued)

Table 4. (Continued)

Measure	Reference level	Fixed effect	Estimate	SE	t	p	R ² m	R ² c	Random effect		
									Factor	Variance	SD
Full word (N = 56)	Period 2	Period 5	0.32	0.12	2.63	.009 ^{***}	.03	.32	Participant	0.28	0.53
									Prompt	0.01	0.12
	Period 3	Period 5	0.47	0.12	4.03	< .001 ^{***}	.03	.32	Participant	0.28	0.53
									Prompt	0.01	0.12
Full word (N = 56)	Period 4	Period 5	0.34	0.12	2.97	.003 ^{**}	.03	.32	Participant	0.28	0.53
									Prompt	0.01	0.12
	Period 2	Period 5	0.39	0.14	2.81	.005 ^{**}	.02	.17	Participant	0.14	0.37
									Prompt	0.02	0.12
Below clause (N = 56)	Period 3	Period 5	0.35	0.13	2.71	.007 ^{**}	.02	.17	Participant	0.14	0.37
									Prompt	0.02	0.12
	Period 2	Period 5	0.46	0.15	3.19	.002 ^{**}	.03	.09	Participant	0.06	0.25
By context											
Precontextual (N = 56)	Period 3	Period 5	-0.42	0.13	-3.23	.001 ^{***}	.03	.24	Participant	0.22	0.47
Contextual (N = 56)	Period 2	Period 5	0.55	0.13	4.29	< .001 ^{***}	.07	.23	Participant	0.16	0.40
	Period 3	Period 4	0.32	0.12	2.62	.009 ^{**}	.07	.23	Participant	0.16	0.40
		Period 5	0.72	0.12	5.93	< .001 ^{***}					
	Period 4	Period 3	-0.32	0.12	-2.62	.009 ^{**}	.07	.23	Participant	0.16	0.40
		Period 5	0.40	0.12	3.29	.001 ^{***}					

p < .01; *p < .001.

Table 5. Significant results from linear mixed-effects models examining the interaction effects between writing period and task complexity on pausing behaviors by individual participants.

Measure	Reference level	Fixed effect	Estimate	SE	t	p	R ² _m	R ² _c	Random effect		
									Factor	Variance	SD
Pause length (in milliseconds, ln)											
Between subsentences (N = 53)											
	Period 3	Period 5: Task	1.10	0.38	2.87	.005**	.07	.12	Participant	0.03	0.18
									Prompt	0.03	0.16
Pause frequency (number of pauses per 100 keystrokes, ln)											
Between sentences (N = 39)											
	Period 2	Period 3: Task	-2.07	0.61	-3.37	.001**	.17	.43	Participant	0.27	0.52
		Period 4: Task	-1.69	0.59	-2.88	.005**					
		Period 5: Task	-1.82	0.59	-3.07	.003**					

**p < .01.

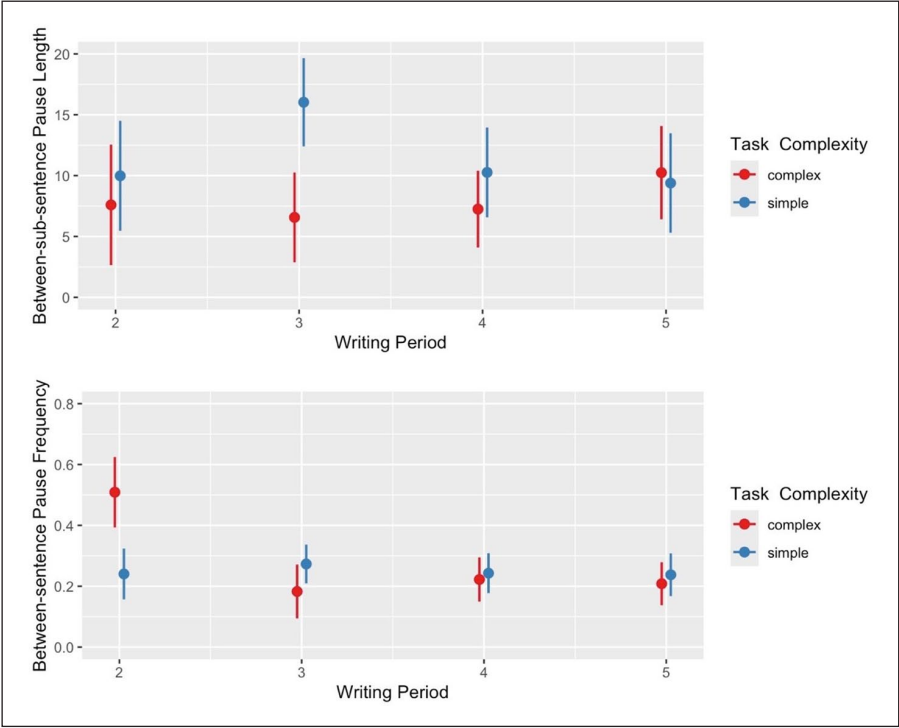


Figure 3. Illustration of significant interaction effects (based on raw data).

writing period accounting for 35% of the variance. However, it is important to interpret this analysis cautiously, as pauses at larger boundaries were produced by only a few participants in each writing period.

Next, we investigated the extent to which writing period differently influenced participants’ cognitive activities underlying the pausing and revision behaviors under each task condition. Tables S18 and S19 in the online supplementary file present the summary of pausing-related stimulated recall comments across the two task versions. Participants reported slightly higher percentage of activities associated with planning and translation in the middle periods under the simple task condition, whereas translation-related activities were referenced more frequently in the last period under the complex task version. Tables S20–S23 in the online supplemental material summarize the stimulated recall results for revision behaviors categorized by linguistic domain and context across the two task versions, respectively. The patterns were largely similar for the two task versions. However, a notable difference was that a higher percentage of revisions was attributed to task requirements in Period 5 in the simple task version (9%) than in the complex task version (1%).

Table 6. Significant results from linear mixed-effects models examining the effects of period on pausing behaviors by individual participants in the complex task version where interactions were significant.

Measure	Reference level	Fixed effect	Estimate	SE	t	p	R ² m	R ² c	Random effect	
									Factor	Variance
Pause frequency (number of pauses per 100 keystrokes, ln)										
Between sentences (N = 30)										
Period 2	Period 3	Period 3	-1.90	0.47	-4.08	< .001***	.35	.51	Participant	0.16
		Period 4	-1.65	0.43	-3.83	< .001***				
		Period 5	-1.79	0.43	-4.16	< .001***				0.40

***p < .001.

VI Discussion

I Temporal distribution of writing behaviors and cognitive activities

Our first research question aimed to understand how the progression of writing over time affects writing behaviors and cognitive activities in computer-assisted collaborative writing. Specifically, it explored whether different writing behaviors, including speed fluency, pausing, and revision, as well as the underlying cognitive activities associated with these behaviors, varied across different writing periods. Findings from a series of linear mixed-effects models indicated that while writing period did not significantly influence speed fluency measures, it did have significant effects on several aspects of pausing and revision behaviors. Participants tended to pause for longer durations within words during Period 5 and more frequently within words in Period 2. Moreover, they engaged in more-frequent revisions during Period 5 compared with earlier periods. Stimulated recall comments further revealed that participants' revisions in Period 5 were more often linked to task requirement-related processes than in previous writing periods.

The findings regarding speed fluency in our study diverge from prior research on individual writing (Michel et al., 2020; Van Waes & Leijten, 2015). Previous studies usually observed a fluctuation in speed fluency between the first and/or last period(s) and middle periods, whereas our participants did not show significant variation in speed fluency across writing periods. This discrepancy in results can potentially be explained by the distinct characteristics of collaborative versus individual writing processes. In collaborative writing contexts, the joint effort in producing a text may alleviate time pressure. This could allow participants more time to engage in collaborative planning of text organization and content during the prewriting stage. As a result, writers may experience a relatively smooth writing process across all periods, without significant fluctuations in text production speed. An alternative or supplementary explanation could be that in collaborative writing settings, participants may leverage linguistic and nonlinguistic resources available to their partners who are typing. They may also alternate in writing turns, which could contribute to maintaining a stable rate of text production over time.

Our findings for pause length and frequency diverge from previous research focused on the temporal dynamics of individual writing. Unlike studies by Barkaoui (2019), Michel et al. (2020), Révész et al. (2023), and Xu and Qi (2017), we found longer pauses within words in Period 5 and more-frequent pauses within words in Period 2. These discrepancies may be attributed to the unique characteristics of collaborative integrated writing settings. The collaborative nature of the task may reduce the time pressure, enabling participants to engage more in notetaking and outlining before actual writing begins. Therefore, in Period 2, participants may have allocated more cognitive resources to microlevel translation processes, such as spelling and linguistic forms, which are reflected in more-frequent pauses within words (Torres, 2023). In Period 5, when participants neared the completion of text drafting, the reduced demand for planning and translation might have shifted their focus to monitoring for orthographic errors, leading to longer pauses within words as they reviewed and revised their writing. This reasoning has been evidenced in the stimulated recall data exemplified in Table 7.

Table 7. Examples of participants’ cognitive activities related to within-word pausing in Periods 2 and 5 (translated).

Example 1 (translation-lexical retrieval, within-word)
Participant I029Q: I was thinking whether to use “human,” “humans,” or “human beings,” as I couldn’t decide between these options.
Example 2 (translation-lexical retrieval, within-word)
Participant I104D: I was thinking about how to spell “customer.”
Example 3 (translation-lexical retrieval, within-word)
Participant I104E: I was thinking whether I can use the word “controversial.”

Table 8. Examples of participants’ cognitive activities related to revision for task requirements in Period 5 (translated).

Example 1 (below-clause, contextual)
Participant I029F: Because the word count exceeded the limit. The requirement specifies 250 words, with a 10% margin. The maximum is actually 275, but it’s already over 300, so I trimmed it down to get closer to the required word count.
Example 2 (full-word, contextual)
Participant I029Y: The original text uses “vulnerable to,” but I didn’t want to use the same phrase since I couldn’t copy directly from the original text.
Example 3 (below-clause, pre-contextual)
Participant I103B: This sentence was subjective, and the requirements stated that personal opinions shouldn’t be included.

Regarding revision behaviors, we observed a gradual increase in revision frequency from earlier periods to Period 5, which was consistent across revisions at various linguistic levels: below-word, full-word, and below-clause revisions. These findings align closely with previous studies by Révész et al. (2023) and Roca de Larios et al. (2008), which also found a higher frequency of revisions occurring in later than earlier writing periods. Moreover, the increasing frequency of contextual revisions over time observed in our study mirrors findings by Barkaoui (2016) and Lu and Révész (2021). In our timed collaborative writing tasks, participants appeared to prioritize drafting a complete text first, which enabled them to do more extensive revisions to already written portions of text toward the end of the writing process with a view to refining and improving the existing content. Such temporal patterns might also reflect participants’ coordination of the cognitive resources during the writing process. The focus on formulation and execution during the drafting process probably limited attentional resources for monitoring. Therefore, participants could only redirect the attention to monitoring when the cognitive resources were alleviated in the last period. This interpretation is further supported by the stimulated recall data. The examples listed in Table 8 demonstrated participants’ revisions made to meet task requirements, such as keep word limit and avoid direct copying and personal opinions, in the last period.

2 Task complexity and temporal distribution of writing behaviors and cognitive activities

The second research question aimed to explore the interaction effect between writing period and task complexity on the temporal distribution of writing behaviors and cognitive activities. Out of the 19 measures examined, we identified significant interaction effects for two indices. Specifically, participants paused longer between sub-sentences in Period 3 than in Period 5 in the simple task version, whereas participants paused longer between subsentences in Period 5 compared with Period 3 in the complex task version. In addition, we observed significantly more-frequent pauses between sentences during Period 2 than other periods in the complex task version. The stimulated recall data also indicated that some difference in the temporal distribution of cognitive writing activities across the two task versions. In the simple task version, a higher percentage of planning and translation activities occurred in the middle writing periods, whereas the last period in the complex task version witnessed more translation activities than previous ones. In addition, revisions made to meet task requirements were more prevalent in the last period of the simple task version.

Our finding that speed fluency was consistent during the writing process in the simple and complex task conditions aligns with much of the previous literature. For example, studies by Jung (2020) and Révész et al. (2017) found similar speed fluency regardless of task complexity. In contrast, participants in Abdi Tabari et al.'s (2024) study exhibited greater speed fluency when completing the simple task. However, it should be noted that the three previous studies only examined speed fluency between the simple and complex task conditions for the whole writing period rather than across writing periods. The lack of interaction effects for speed fluency in our study, as previously discussed, might be attributed to the nature of integrated tasks. This task type probably allowed writers to plan the entire text before writing regardless of the task complexity manipulation. In addition, the collaborative aspect of the tasks might have enabled participants type in turn, thereby avoiding writing breakdowns and maintaining stable speed fluency over time across both task versions.

Turning to the temporal distribution of pausing behaviors across tasks of varying complexity, the two interaction effects likely arise from differences in how participants regulated their progress across the two task versions. The complex task version required greater initial time for reading, comparing, and contrasting information across the three source texts. Participants also needed additional time for planning to organize the structure of their texts. As a result, most participants just began actively composing during Period 2 under the complex task condition. This intensive planning at the start of the writing process increased the frequency of between-sentence pauses during Period 2. Furthermore, the delayed initiation of writing in the complex task condition caused many participants to still be drafting during Period 5. The cognitive demands of completing text production under time constraints likely contributed to longer subsentence pauses during this phase. Recall data also corroborated this trend, as translation activities slightly increased in the last period. In contrast, the simple task condition, which required less information integration, enabled participants to begin writing earlier, engage more deeply in cognitive processes during the middle periods, and finish text production sooner. As a result, the simple task condition did not exhibit significant changes in

between-sentence pause frequency and experienced longer subsentence pauses during the middle phases of writing.

Our study did not find any significant influence of task complexity on the temporal distribution of revision behaviors, contrary to findings from previous studies examining task complexity effects on the whole writing process in individual writing (Jung, 2020; Révész et al., 2017). As discussed previously, the collaborative process of joint notetaking and outlining before composing likely contributed to a clearer initial plan, which may have minimized the need for substantial content revisions during the execution and monitoring phases of writing in either task condition. In addition, the collaborative nature of our writing task may have reduced time pressure and thus allowed for more thorough online planning, thereby potentially decreasing the need for extensive revisions regardless of task complexity. Furthermore, collaborative writing environments often prompt discussion about linguistic forms (Storch, 2021), which might have preemptively addressed language-related issues during both task versions, thereby reducing the necessity for revisions focused on language refinement. However, the stimulated recall indicated that revisions made to meet task requirements in Period 5 almost exclusively occurred in the simple task condition. This also suggested that participants were able to allocate more cognitive resources to monitoring in the final period while performing the simple task version.

3 Limitations and future research

Before drawing conclusions from our study, it is important to acknowledge several limitations that should be considered in interpreting the findings. First, we did not account for potential variability in collaborative patterns among pairs. While previous research suggests that interactive patterns do not significantly change with task complexity in collaborative writing (Hsu, 2020), future studies could explore how task complexity influences collaborative patterns in synchronous settings and how these interactions may affect writing behaviors and the associated cognitive activities over time. Second, although the source texts used were made to be similar in linguistic complexity, we did find differences in topic and content familiarity. Future research could investigate the influence of topic familiarity on the temporal distribution of writing processes. When it comes to data analyses, our study followed the convention in the literature (e.g., Wengelin, 2006) by employing a pause threshold of 2 seconds. Using a lower pause threshold (e.g., 200 ms) in future research could provide a more detailed understanding of pausing behaviors in collaborative writing, capturing more subtle cognitive processes. In addition, nonlinear pauses, which typically involve peer interaction and role shifts in typing, were excluded from the analysis due to the considerable variability in their duration. Future research could incorporate pair talk alongside pause analysis to examine learners' writing processes during non-linear pausing. Moreover, this study did not include data on pair talk. Further triangulation of this type of data would yield more insights into the collaborative writing process. Another limitation is a lack of information on participants' viewing behavior during writing. Reading source texts and previously produced text is also a key writing process during integrated writing (Michel et al., 2020), which we did not capture in the current study. Future research could benefit from triangulating eye-tracking with keystroke logging and verbal protocols to obtain a

comprehensive understanding of the temporal dynamics of writing behaviors and cognitive activities.

VII Conclusion

In this study, our goal was to investigate how L2 writers' speed fluency, pausing, revision behaviors, and underlying cognitive activities varied across five different writing periods in computer-assisted collaborative writing tasks. In addition, we have explored the influence of task complexity on the temporal distribution of these writing behaviors and cognitive activities. Our research yielded two main findings. First, participants exhibited longer pauses within words and higher revision frequency in the last writing period, whereas they paused more frequently within words in the second period. Second, participants in the complex task version paused more frequently between sentences in the second period than later periods. From a methodological standpoint, our study highlighted the usefulness of combining keystroke logging with stimulated recall to capture the temporal evolution of learners' writing processes in computer-assisted collaborative tasks.

Last but not least, our findings have some key pedagogical implications. Based on our findings, it would appear that novice academic writers may benefit from engaging in collaborative writing tasks. The opportunity to collaborate seems to lead to reduced cognitive load, allowing writers to engage in smoother and more fluent text production. Our results also suggest that learners may be less affected by timing and task complexity differences when engaged in collaborative writing. In turn, collaborative writing tasks may have the capacity to push learners to complete tasks that might otherwise be too demanding to tackle on their own due to constraints imposed by time or cognitive task demands. Considering that most of our participants were not familiar with collaborative writing, our study provided evidence in support of promoting collaborative writing practice in academic writing settings.

Acknowledgements

We would like to thank the handling editor and anonymous reviewers for their comments on the previous versions of our manuscript. We are grateful to all participants involved in our study. In addition, special thanks go to the raters and coders for their assistance in the rating and coding work.

Data availability

Data will be made available upon request.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was funded by The International Research Foundation for English Language Education (TIRF) and a doctoral scholarship awarded to the first author by China Scholarship Council (number 202108060053).

Ethical approval and informed consent statement

This study has been approved by University College London in March 2022. Written consent was obtained from participants before data collection.

ORCID iD

Xin Rong  <https://orcid.org/0009-0005-1922-771X>

Supplemental material

Supplemental material for this article is available online.

References

- Abdi Tabari, M., Abdel Latif, M.M.M., & Tian, Y. (2024). The impact of computer-mediated task complexity on writing fluency: A comparative study of L1 and L2 writers' fluency performance. *Computers and Composition*, 73, Article 102863. <https://doi.org/10.1016/j.comp-com.2024.102863>
- Baaijen, V.M., Galbraith, D., & de Glopper, K. (2012). Keystroke analysis: Reflections on procedures and measures. *Written Communication*, 29(3), 246–277. <https://doi.org/10.1177/0741088312451108>
- 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(1), 320–340.
- Barkaoui, K. (2019). What can L2 writers' pausing behavior tell us about their L2 writing processes? *Studies in Second Language Acquisition*, 41(3), 529–554. <https://doi.org/10.1017/S027226311900010X>
- Barton, K. (2020). *MuMIn: Multi-model inference. R package (Version 1.47.5)*. <https://CRAN.R-project.org/package=MuMIn>
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Chen, W., & Hapgood, S. (2021). Understanding knowledge, participation and learning in L2 collaborative writing: A metacognitive theory perspective. *Language Teaching Research*, 25(2), 256–281. <https://doi.org/10.1177/1362168819837560>
- Elabdali, R. (2021). Are two heads really better than one? A meta-analysis of the L2 learning benefits of collaborative writing. *Journal of Second Language Writing*, 52, Article 100788. <https://doi.org/10.1016/j.jslw.2020.100788>
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39(2), 175–191. <https://doi.org/10.3758/BF03193146>
- 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. <https://doi.org/10.1111/lang.12280>
- Gass, S.M., & Mackey, A. (2017). *Stimulated recall methodology in applied linguistics and L2 research* (2nd ed.). Routledge.
- Hall, S., Baaijen, V.M., & Galbraith, D. (2024). Constructing theoretically informed measures of pause duration in experimentally manipulated writing. *Reading and Writing*, 37, 329–357. <https://doi.org/10.1007/s11145-022-10284-4>
- Hayes, J.R. (2012). Modeling and remodeling writing. *Written Communication*, 29(3), 369–388.
- Hsu, H.-C. (2020). The impact of task complexity on patterns of interaction during web-based asynchronous collaborative writing tasks. *System*, 93, Article 102328. <https://doi.org/10.1016/j.system.2020.102328>
- Jung, J. (2020). Effects of content support on integrated reading-writing task performance and incidental vocabulary learning. *System*, 93, Article 102288. <https://doi.org/10.1016/j.system.2020.102288>

- Kellogg, R.T. (1996). A model of working memory in writing. In C.M. Levy & S. Ransdell (Eds.), *The science of writing: Theories, methods, individual differences and applications* (pp. 57–71). Lawrence Erlbaum Associates.
- Koo, T.K., & Li, M.Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, 15(2), 155–163. <https://doi.org/10.1016/j.jcm.2016.02.012>
- Kuznetsova, A., Brockhoff, P.B., & Christensen, R.H.B. (2017). LmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, 82(13), 1–26. <https://doi.org/10.18637/jss.v082.i13>
- Leijten, M., & Van Waes, L. (2013). Keystroke logging in writing research: Using Inputlog to analyze and visualize writing processes. *Written Communication*, 30(3), 358–392. <https://doi.org/10.1177/0741088313491692>
- Leow, R. (2015). *Explicit learning in the classroom A student-centered perspective*. Routledge.
- Lindgren, E., & Sullivan, K.P.H. (2006). Analysing online revision. In K.P.H. Sullivan & E. Lindgren (Eds.), *Computer key-stroke logging and writing* (pp. 157–188). Elsevier.
- 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, Article 102544. <https://doi.org/10.1016/j.system.2021.102544>
- Lüdecke, D. (2020). *sjPlot: Data visualization for statistics in social science. R package (Version 2.8.4)*. <https://CRAN.R-project.org/package=sjPlot>
- Mak, B., & Coniam, D. (2008). Using wikis to enhance and develop writing skills among secondary school students in Hong Kong. *System*, 36(3), 437–455. <https://doi.org/10.1016/j.system.2008.02.004>
- Manchón, R.M., & Roca de Larios, J. (2007). On the temporal nature of planning in L1 and L2 composing. *Language Learning*, 57(4), 549–593. <https://doi.org/10.1111/j.1467-9922.2007.00428.x>
- McDonough, K., Crawford, W., & De Vleeschauwer, J. (2016). Thai EFL learners' interaction during collaborative writing tasks and its relationship to text quality. In M. Sato & S. Ballinger (Eds.), *Peer interaction and second language learning: Pedagogical potential and research agenda* (pp. 185–208). John Benjamins.
- Michel, M., Révész, A., Lu, X., Kourтали, N., Lee, M., & Borges, L. (2020). Investigating L2 writing processes across independent and integrated tasks: A mixed-methods study. *Second Language Research*, 36(3), 307–334. <https://doi.org/10.1177/0267658320915501>
- Orne, M.T. (1962). On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist*, 17(11), 776–783. <https://doi.org/10.1037/h0043424>
- R Core Team. (2023). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Roca de Larios, J., Manchón, R., Murphy, L., & Marín, J. (2008). The foreign language writer's strategic behaviour in the allocation of time to writing processes. *Journal of Second Language Writing*, 17(1), 30–47. <https://doi.org/10.1016/j.jslw.2007.08.005>
- Roca de Larios, J., Marín, J., & Murphy, L. (2001). A temporal analysis of formulation processes in L1 and L2 writing. *Language Learning*, 51(3), 497–538. <https://doi.org/10.1111/0023-8333.00163>
- Révész, A. (2014). Towards a fuller assessment of cognitive models of task-based learning: Investigating task-generated cognitive demands and processes. *Applied Linguistics*, 35(1), 87–92. <https://doi.org/10.1093/applin/amt039>
- Révész, A., Kourтали, N., & Mazgutova, D. (2017). Effects of task complexity on L2 writing behaviors and linguistic complexity. *Language Learning*, 67(1), 208–241. <https://doi.org/10.1111/lang.12205>

- Révész, A., & Michel, M. (2019). Introduction. *Studies in Second Language Acquisition*, 41(3), 491–501. <https://doi.org/10.1017/S0272263119000329>
- Révész, A., Michel, M., & Lee, M. (2019). Exploring second language writers' pausing and revision behaviors: A mixed-methods study. *Studies in Second Language Acquisition*, 41(3), 605–631. <https://doi.org/10.1017/S027226311900024X>
- Révész, A., Michel, M., & Lee, M. (2023). Exploring the relationship of working memory to the temporal distribution of pausing and revision behaviors during L2 writing. *Studies in Second Language Acquisition*, 45(S3), 680–709. <https://doi.org/10.1017/S0272263123000074>
- Rijlaarsdam, G., & van den Bergh, H. (1996). The dynamics of composing — an agenda for research into an interactive compensatory model of writing: Many questions, some answers. In C.M. Levy & S. Ransdell (Eds.), *The science of writing: Theories, methods, individual differences, and applications* (pp. 107–125). Lawrence Erlbaum Associates.
- Rijlaarsdam, G., & van den Bergh, H. (2006). Writing process theory: A functional dynamic approach. In C.A. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research* (pp. 41–53). The Guilford Press.
- Robinson, P. (2001). Task complexity, task difficulty, and task production: Exploring interactions in a componential framework. *Applied Linguistics*, 22(1), 27–57. <https://doi.org/10.1093/applin/22.1.27>
- Rong, X., & Révész, A. (2025). The effects of task complexity on second language writing processes and products in computer-assisted collaborative writing. *Journal of Second Language Writing*, 69, Article 101216, <https://doi.org/10.1016/j.jslw.2025.101216>
- Rouhshad, A., Wigglesworth, G., & Storch, N. (2016). The nature of negotiations in face-to-face versus computer-mediated communication in pair interactions. *Language Teaching Research*, 20(4), 514–534. <https://doi.org/10.1177/1362168815584455>
- 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(3), 201–233. <https://doi.org/10.1016/j.jslw.2006.06.002>
- Storch, N. (2013). *Collaborative writing in L2 classrooms*. Multilingual Matters.
- Storch, N. (2016). Collaborative writing. In R. M. Manchón & P. K. Matsuda (Eds.), *Handbook of second and foreign language writing* (pp. 387–406). De Gruyter.
- Storch, N. (2019). Collaborative writing. *Language Teaching*, 52(1), 40–59. <https://doi.org/10.1017/S0261444818000320>
- Storch, N. (2021). Collaborative writing: Promoting languaging among language learners. In M. García Mayo (Ed.), *Working collaboratively in second/foreign language learning* (pp. 13–34). De Gruyter Mouton.
- Teng, M.F., & Huang, J. (2021). The effects of incorporating metacognitive strategies instruction into collaborative writing on writing complexity, accuracy, and fluency. *Asia Pacific Journal of Education*, 43(4), 1071–1090. <https://doi.org/10.1080/02188791.2021.1982675>
- Tillema, M. (2012). Writing in first and second language: Empirical studies on text quality and writing processes. *Unpublished doctoral dissertation, Netherlands Graduate School of Linguistics*. https://www.lotpublications.nl/Documents/299_fulltext.pdf
- Tolmie, A., Muijs, D., & McAteer, E. (2011). *Quantitative methods in educational and social research using SPSS*. Open University Press.
- Torres, J. (2023). Exploring working memory and language dominance in heritage bilinguals' writing processes. *Studies in Second Language Acquisition*, 45(3), 710–736. <https://doi.org/10.1017/S027226312300013X>
- 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>

- Van Weijen, D. (2009). Writing processes, text quality, and tasks effects: Empirical studies in first and second language writing. Unpublished doctoral dissertation, Netherlands Graduate School of Linguistics. https://www.lotpublications.nl/Documents/201_fulltext.pdf
- Wengelin, A. (2006). Examining pauses in writing: Theory, methods and empirical data. In K.H. Sullivan & E. Lindgren (Eds.), *Computer keystroke logging and writing: Methods and applications* (pp. 107–130). Elsevier Science.
- Wigglesworth, G., & Storch, N. (2009). Pair versus individual writing: Effects on fluency, complexity and accuracy. *Language Testing*, 26(3), 445–466. <https://doi.org/10.1177/0265532209104670>
- Xu, C., & Qi, Y. (2017). Analyzing pauses in computer-assisted EFL writing—a computer-key-stroke-log perspective. *Educational Technology & Society*, 20(4), 24–34.