Feedback Research in Second Language **EUROKD**

https://doi.org/10.32038/frsl.2025.03.02

From pauses to processes: Automating the classification of L2 writing processes from keystroke logs

Aitor Garcés-Manzanera
University of Murcia, Spain

Correspondence
Email: aitor.garces@um.es

Abstract

Although cognitive models of writing have long emphasized the importance of planning, formulation, and revision, little is known about how adult L2 writers distribute these subprocesses during real-time composition. This study addresses that gap by investigating the temporal and spatial characteristics of writing behaviors in a timed argumentative task. Sixteen Spanish university students participated in a 35-minute English writing task, recorded with Inputlog. Writing events were automatically classified using a VBA-based system that integrates pause duration, pause location, and event context to identify underlying subprocesses. The analysis examined how these subprocesses evolved over time, how they related to linguistic pause locations, and how pausing behavior differed across subprocesses when measured as pauses per minute, mean pause duration, and relative pause frequency. Results indicated that subprocess distribution remained relatively stable throughout the session, but clear associations emerged between pause location and subprocess type, for instance, within-word and before-word pauses were linked to formulation, while sentence- and paragraph-initial pauses were associated with planning. Differences in pause duration and frequency also reflected distinct cognitive demands, with planning marked by longer pauses and revision by shorter, more frequent ones. The results support the classification system and shed light on cognitive effort in L2 writing, with potential applications for feedback and task design.

ARTICLE HISTORY

Received: 12 April 2025
Revised: 22 July 2025
Accepted: 02 August 2025

KEYWORDS

L2 writing, keystroke logging, writing processes, cognitive effort, pause analysis

Introduction

Writing is a key skill in both academic and professional settings, and a central indicator of advanced L2 competence. Earlier studies tended to focus on written outcomes such as accuracy, cohesion or overall text quality. In the past decade, there has been increasing interest in examining how writing develops during the act itself (Manchón, 2011; Sasaki et al., 2018). This change has been supported by the spread of digital tools that let researchers observe writing in progress. The advent of keystroke logging software (KLS), in particular, has allowed for close examination of planning, formulation, and revision during the act of writing itself (Leijten & Van Waes, 2013; Révész et al., 2023).

Among the most commonly studied indicators of cognitive activity in writing are pauses, understood as short periods of inactivity that may signal transitions between subprocesses. When interpreted in context, they can shed light on how writers handle lexical retrieval, organise content, or carry out local and global revisions (Medimorec & Risko, 2017; Hall et al., 2024). However, many studies still rely on fixed pause thresholds or treat pauses as isolated markers, without relating them systematically to the different subprocesses involved.

This study addresses that gap by analysing how adult L2 writers allocate time to planning, formulation, and revision, including their respective subcomponents, across the writing task. Using an automated system that classifies keystroke events based on pause duration, structural location, and contextual information, this study investigates the temporal distribution of writing processes, the association between pause location and subprocess type, and the frequency and duration of pausing behavior across subprocesses.

Theoretical Framework

Writing as a Cognitively Demanding Activity

Cognitive models of writing have long conceptualised text production as a complex and dynamic activity involving the continuous interplay of working memory, long-term memory, and executive control. In their foundational work, Flower and Hayes (1981) proposed that writing involves three key components: the task environment, the writer's long-term memory, and a set of cognitive processes responsible for generating and shaping text. These processes - planning, translating, and reviewing - operate recursively and are monitored by a central executive mechanism. This recursive view replaced earlier linear conceptions and foregrounded the cognitive demands of composing, particularly the role of attentional control and memory in managing competing demands during writing.

In his revised versions of the model (1996, 2012), Hayes incorporated motivational states, affective factors, and contextual influences, drawing also on evidence from studies in digital writing. The 2012 update gave more weight to executive control and the writer's goals, acknowledging that task demands and the chosen medium can shape how cognitive processes unfold. Similarly, Kellogg (1996) proposed a model based on Baddeley's theory of working memory, assigning specific roles to each of its components. The phonological loop was linked to verbal repetition and lexical access, while the visuospatial sketchpad was seen as responsible for managing content structure and overseeing the visual flow of the text. Under conditions of increased linguistic or conceptual difficulty, cognitive load on these subsystems may trigger disfluencies, such as hesitations or revisions.

Along similar lines, Galbraith (2009) described writing as the result of two types of processing. Some language emerges with little effort, activated through familiar patterns stored in memory. In parallel, writers sometimes monitor what they produce more deliberately. They review, adjust, or reorganise the text when it does not yet reflect what they aim to communicate. This helps account for why some sentences appear fluent and continuous, whereas others are interrupted, slower, or revised several times. These two modes can overlap within the same task, which raises the need for research tools capable of tracing how they alternate over time.

While theoretical models place emphasis on different elements of writing, there is broad agreement that writing does not unfold in a fixed, linear fashion. Writers may pause, rethink earlier decisions, or return to parts already drafted. These actions are not random; they tend to respond to how the task evolves, to the content being developed, or to doubts that arise mid-process. Previous research has shown that what are usually described as separate phases (i.e., planning, composing, and revising) often appear in a flexible, interwoven way, depending on what the situation demands (Roca de Larios et al., 2008; Tillema, 2012; Michel et al., 2020).

Importantly, these models do not merely describe writing at a general level, but also identify specific subprocesses that can be empirically distinguished. Planning, for instance, can be subdivided into global planning - concerned with macro-level textual organization - and local planning, involving immediate structuring decisions. Formulation may range from fluent production to problem-solving behaviors during lexical or syntactic encoding. Revision can be classified as subword (e.g., letter or orthographic changes), word-level (e.g., lexical substitutions), or above-word (e.g., sentence reorganization). This level of granularity, described both in theoretical accounts and operationalised in empirical work (see Kellogg, 1996; Garcés-Manzanera, 2021, in children's L2 writing), provides the necessary foundation for process-tracing methodologies that attempt to classify subprocesses on the basis of observable traces, such as keystrokes and pauses.

These theoretical distinctions support the use of keystroke logging for identifying subprocesses in real time. Yet, to interpret such data meaningfully, a pause must be situated within its local context - textual, positional, and behavioral. The next section addresses how pauses have been treated in writing research, and how empirical developments have moved from simplistic threshold-based models to more integrated and cognitively interpretable systems.

Pauses as Traces of Writing Processes

Pauses, understood as periods of inactivity between keystrokes, are commonly treated as surface-level manifestations of underlying cognitive processes. Empirical studies have consistently reported that longer pauses tend to coincide with cognitively demanding operations, including macro-level planning, lexical search, and revision (Chenoweth & Hayes, 2001; Medimorec & Risko, 2017). Still, neither pause length nor its textual position can, on their own, clarify the nature of the mental process underway.

Initial research efforts often relied on fixed time thresholds to flag cognitively meaningful pauses. Durations exceeding 2000 milliseconds, for example, have frequently been interpreted as signs of planning or processing difficulty (Barkaoui, 2019; Révész et al., 2019). However, such interpretations have drawn criticism. Fixed thresholds show poor adaptability across writers and tasks, and their contextual relevance remains questionable (Wengelin, 2006; Barkaoui, 2019). A pause of 2500 milliseconds placed between two independent clauses may suggest conceptual reorganization, while a pause of identical length within a single word may reflect difficulties in lexical retrieval or spelling.

More recent studies recommend a shift away from purely temporal criteria. Hall et al. (2022), for example, note that pause length gains interpretive value only when linked to the type of cognitive shift involved, such as switching from text generation to monitoring or revision. Their results question the utility of duration-based thresholds as stand-alone indicators. Conijn et al. (2022) similarly argue that combining pause data with information on nearby keystroke events and revision behavior yields more robust inferences. These findings converge in supporting classification methods that incorporate both timing and local textual context to interpret writing behavior more accurately.

Some studies have explored the use of statistical modelling techniques, such as mixture models, to disentangle overlapping pause patterns. Such empirical endeavors have shown that pauses are not normally distributed but tend to form several clusters, each associated with a different type of cognitive or mechanical activity (Van Waes & Leijten, 2015; Conijn et al., 2021). With these studies, researchers have thus far been able to distinguish brief motor delays from longer pauses, which are more likely to reflect planning, retrieval or other processing demands.

Although these methods have advanced the analysis of writing processes, their application across studies remains uneven. In many cases, studies have relied on fixed duration thresholds or describe pauses in general terms (as noted by Hall et al., 2024 and Conijn et al., 2025). Another point worth mentioning is that few of these procedures have been incorporated into automated tools designed to handle large-scale writing data, which has thus limited their potential for broader empirical use.

While these advances have clarified some methodological constraints in pause-based research, their application remains inconsistent and often disconnected from cognitive models of writing. Many studies continue to rely on static thresholds or provide general descriptions of pauses without establishing a principled correspondence with specific subprocesses. Automated systems capable of handling large datasets rarely incorporate classification rules grounded in theory. This issue was already raised in Abdel Latif's (2008) review of real-time writing research and remains unresolved. More recent work points to the same problem: Hall et al. (2022) note the absence of agreed criteria to distinguish between cognitive and non-cognitive pauses, while Conijn et al. (2025) acknowledge the difficulty of segmenting keystroke data reliably, particularly when annotating transitions between phases. This body of evidence highlights the need for a classification system that goes beyond surface description and instead applies theoretically grounded distributions to pause analysis. Such an approach avoids reliance on arbitrary fixed thresholds and offers a clearer picture of the cognitive activities reflected in pause durations.

To address this gap, the present study adopts a pause classification perspective that integrates temporal information (duration), structural location (e.g., within-word, between-word, between-sentence), and local context (type of keystroke event) to infer subprocesses. The system moves beyond conventional threshold-based criteria by incorporating distinctions drawn from cognitive theory and supported by empirical evidence.

Inputlog and Keystroke Logging Software

Inputlog is a keystroke logging tool designed specifically for writing research as it captures all keyboard and mouse activity during writing, assigning precise time stamps to each event (Leijten & Van Waes, 2013). Beyond recording the temporal sequence of production, it also registers spatial elements such as cursor movements and the specific location of textual changes. These features make it particularly suitable for tracking real-time writing behaviors.

Evidence from empirical work confirms that Inputlog can reveal patterns linked to cognitive effort. Révész et al. (2023) report that pause length and position, as recorded with Inputlog, relate to working memory constraints and the shifting cognitive demands of different writing phases. Longer interruptions tended to occur at clause boundaries or during periods of planning and problem resolution, while brief pauses were more typical of fluent production.

Despite its analytic potential, Inputlog has also prompted methodological questions. Hall et al. (2022) highlight the lack of consistency across studies in how pauses are classified, with many failing to explain or justify the thresholds used. Conijn et al. (2022) provide a refined taxonomy of revision types, yet their focus remains on describing visible modifications rather than interpreting the mental operations behind them. Several scholars have questioned whether observable features like pauses can be reliably linked to underlying subprocesses. Tian and Cushing (2025) caution that a pause might reflect genuine planning, but could equally result from disengagement or environmental distraction. As Schilperoord (2001) notes (cited in Hall et al., 2024), during major restructurings of a text, measurable features such as pause length may not align with distinct cognitive processes. Disagreements over pause thresholds and uncertainty surrounding the function of insertions or revisions complicate the interpretability of such data (Hall et al., 2024; Baaijen et al., 2012).

To respond to these issues, this study introduces an automated classification system embedded in Inputlog. The system links observed events, such as pauses, insertions, deletions, and textual revisions, to writing subprocesses based on established theoretical models. The basis of its classification rests on decision rules that draw on empirical findings and clearly defined criteria to support consistent and process-aware categorisation.

From Pauses to Processes: A Classification Perspective

Most research using keystroke data has analysed pauses in isolation or treated them as indicators of global fluency. Very few studies have attempted to systematically classify cognitive subprocesses on the basis of pause behavior, with the exception of a study by Garcés-Manzanera (2021) on children's L2 writing. As a result, the potential of keystroke logging to reveal transitions between planning, formulation, and revision remains underexploited.

This study introduces a classification system that draws on cognitive models of writing and incorporates findings from previous empirical research. Rather than relying on surface tags or preset thresholds, the system combines information about pause length, position in the text, and the nature of surrounding actions to assign events to specific subprocesses. It builds on recent methodological work (Garcés-Manzanera, 2025, forthcoming) and proposes categories such as global and local planning, fluent and effortful formulation, and revision at different textual levels, from individual letters to whole sentence segments.

The system does not treat pauses in isolation. Instead, it links each event to a cognitive operation described in prior literature, using a set of decision criteria based on observed patterns. This allows for a more transparent and interpretable analysis, particularly when working with large writing datasets. This multi-metric approach is supported by previous research indicating that combining several metrics from keystroke data may yield more robust inferences (Galbraith & Baaijen, 2019).

The Study

While previous research has described planning, formulation, and revision as core components of the writing process, their operationalisation in keystroke logging studies remains limited. Existing classifications tend to rely on duration thresholds or surface-level behaviors, without systematically linking observable events to cognitively defined subprocesses. As noted in recent work, pause duration alone is insufficient to distinguish between types of processing, and revision tagging systems continue to face challenges in how they are defined and applied, pointing to a need for greater transparency (Hall et al., 2024; Conijn et al., 2024, 2022).

The present study addresses this gap through an automated classification system that identifies writing subprocesses based on pause characteristics, event context, and structural location. The system applies subcategorical distinctions within planning, formulation, and revision, and is tailored to adult L2 writing. Unlike the manual annotation procedures used in previous work with child L2 writers (Garcés-Manzanera, 2021), the present version applies an automated system that incorporates subcategories within each macroprocess and targets the event-level characteristics of adult L2 writing.

The following research questions guide the analysis:
RQ1: How are planning, formulation, and revision processes - along with their respective subcategories - distributed across the writing task in adult L2 writers?
RQ2: Is there a relationship between pause location (e.g., within-word, between-word, between-sentence, between-paragraph) and the type of writing subprocess identified?
RQ3: How frequently do pauses occur within each classified writing process, and to what extent does this frequency differ across subprocesses?

Methods

Context and Participants

This study involved 16 adult university students enrolled in a compulsory English as a Foreign Language (EFL) module (2 hours and 30 minutes per week) as part of the Degree in Primary Education at a Spanish university. Participants were recruited through convenience sampling, and all were native speakers of Spanish. The age range was 18 to 23 years ($M = 19.5$, $SD = 1.97$).

Proficiency levels were determined using the Oxford Placement Test. Scores ranged from B1.1 to C1, with 37.5% classified as B1.1, 25% as B1.2, 25% as B2, and 12.5% as C1. A chi-square goodness-of-fit test indicated no significant deviation from a uniform distribution across CEFR levels, $\chi^2(3) = 2.00$, $p = .57$.

Although participants had regular exposure to writing tasks as part of their EFL instruction, such activities were infrequent and not embedded within a structured writing curriculum. None of the participants reported extensive experience with academic writing in English beyond the classroom.

Writing Task and Data Collection

Participants completed a timed argumentative essay in the school's computer lab during a regularly scheduled English class. The writing prompt was designed to elicit connected discourse and encourage the use of organizational and revision strategies in an ecologically valid setting. The topic, *Education and cultural exchange*, invited students to discuss how education can foster intercultural understanding and international cooperation, with optional sub-prompts on study-abroad programs and the inclusion of global perspectives in national curricula. The minimum required length was 200 words, and the time limit was set at 35 minutes.

This task type was selected based on both pedagogical familiarity and its known capacity to engage multiple writing processes. Argumentative writing had been introduced in prior coursework, albeit sporadically, and the topic was deemed accessible given the participants' academic background. From a research standpoint, argumentative essays are particularly suited for the analysis of cognitive operations in L2 writing. Compared to narrative or descriptive tasks, they have been shown to elicit more frequent and complex revisions, especially under time constraints (Medimorec & Risko, 2017; Vandermeulen et al., 2024). Sentence-level and paragraph-level pauses in such tasks are often associated with increased planning demands and structural decision-making (Spelman Miller, 2000; Révész et al., 2019).

Writing sessions were conducted under exam-like conditions using Microsoft Word. No spellcheckers, online dictionaries, or external resources were allowed. Inputlog 9.0 was used to unobtrusively record keystrokes, mouse movements, and timing information throughout the writing process. The session was supervised by the research team, and students were instructed to write as they normally would in a classroom activity.

Once the writing sessions were completed, the Inputlog data were analysed using a custom-built system developed in VBA (Garcés-Manzanera, 2025, forthcoming). This tool assigned a process label to each logged event, based on a set of categories drawn from cognitive models of writing. Events were classified under planning, formulation, or revision, and further grouped into more specific types within each process. After classification, the data were used to calculate how often each process occurred, how much time writers spent on them, and how their distribution varied over the course of the task. To capture potential changes over time, the writing timeline was divided into three segments: beginning, middle, and end.

The essays analysed in this study were collected as part of a data collection phase coordinated by Dr Raquel Criado within a research project funded by the Spanish State Research Agency (AEI, reference: PID2022-137544NB-I00), whose principal investigators were Dr Lourdes Cerezo and Dr Rosa Manchón.

Classification System

This study employed a custom classification system developed in Visual Basic for Applications (VBA), designed to process the output files generated by Inputlog. The system categorises each recorded writing event into one of three core processes: planning, formulation, or revision. These categories are further broken down into specific subprocesses, drawing on established cognitive models of writing (Hayes, 2012; Kellogg, 1996), empirical research on pausing and revision behavior (Hall et al., 2024; Medimorec & Risko, 2017; Révész et al., 2023), and prior manual annotation procedures (Garcés-Manzanera, 2021).

The keystroke data were retrieved using Inputlog's General Analysis module (Leijten & Van Waes, 2013), which logs all keyboard and mouse actions with time-stamped precision. This module was chosen for its ability to deliver consistent and detailed data across participants, including information about event type (e.g., insertions, deletions, navigations), duration of actions and pauses, cursor movements, and text position. The XML files were automatically parsed to reconstruct the writing sequence and extract contextual features for each event. These data served as the basis for the automated classification system, which assigned process type and subprocess labels to each action using a rule-based system informed by cognitive models of writing.
Each writing event was assigned to a single category based on a combination of pause duration, structural location, keypress sequence, and cursor behavior. The classification rules operated hierarchically and were implemented sequentially within the VBA system. Table 1 presents a summary of the criteria used for each subprocess.

Table 1

Classification Criteria for Writing Subprocesses

Process	Subprocess	Key Features Used for Classification
Planning	Global planning	Pause > 5000 ms; pre-sentence or pre-paragraph; no immediate textual insertion
	Local planning	Pause 2000–5000 ms; sentence-internal position; typically followed by a brief insertion or mouse event
Formulation	Fluent formulation	Pause < 1000 ms; within-word; uninterrupted insertion pattern
	Lexical retrieval	Pause 2000–5000 ms; between-word; followed by insertion of content word
	Problem-solving formulation	Pause > 5000 ms; preceding new sentence or structural reorganization
Revision	Subword revision	Pause < 500 ms; edit ≤ 3 characters; within-word or mid-word correction
	Word-level revision	Pause 500–2000 ms; edit between 4–10 characters; single-word replacement or retyping
	Above-word revision	Pause > 2000 ms; edit > 10 characters or affecting structure beyond word level

The criteria applied in the system were based on previous research showing that certain pause lengths and their position in the text are often linked to different types of cognitive effort (Chenoweth & Hayes, 2001; Barkaoui, 2016). The time windows in Table 1 are decision ranges used by the classifier to assign subprocess labels. They are not global minimum cutoffs for detecting or counting pauses in later analyses. To separate planning from revision, especially in cases where the text was modified after some delay or outside a linear flow, the analysis included cursor movements and deletion patterns. Rather than focusing on isolated pauses or fixed time thresholds, the system relied on the context of each event and how it related to what came before and after.

A key difference between this system and earlier ones lies in how writing events are classified and interpreted. Instead of relying on manual coding or surface indicators, the system applies a set of recursive rules that take into account the context in which each event occurs. For every entry in the log, the system records the assigned process and subprocess, the segment of the task in which it takes place, its location in the document, and any pause associated with it. Once all events have been labelled, the data are compiled to examine how different types of processing are distributed over time, both at the level of broad categories and more specific subprocesses.

The system was developed to overcome several shortcomings found in previous classification methods. A central strength was the integration of cognitive theory into the classification logic, which helped clarify the interpretation of writing behaviors. Also, the design took into account recent criticisms calling for clearer definitions and greater transparency in keystroke logging research (see Hall et al., 2024; Conijn et al., 2022). In light of this, we attempted that the automated nature of the system made the analysis both consistent and reproducible across participants, while still allowing manual verification of classification decisions when necessary.

While the classification system was fully automated, we conducted an internal check: two randomly selected participants were hand coded at the subprocess level and compared with the automated labels. Agreement reached 83.0% (kappa = .727; n = 2,384) and 77.4% (kappa = .673; n = 2,082). Most mismatches occurred where the algorithm labeled planning-local for short boundary pauses that a human rater judged as formulation, and when large insertions or deletions coincided with boundary pauses.

Measures

All measures were derived from the automated classification system described in Section 3.4. Each writing event was labelled for process (planning, formulation, or revision), subprocess (e.g., global planning, fluent formulation, word-level revision), pause duration, pause location, and timestamp. These classifications were used to compute a set of aggregate measures consistent with the study's research questions and grounded in prior writing process research (e.g., Medimorec & Risko, 2017; Révész et al., 2023; Van Waes & Leijten, 2015). In RQ2 and RQ3, the term pause refers to the inter-event delay between logged actions. No lower duration bound was imposed. We keep the word pause for readability while using all delays present in the log.

For **RQ1**, which examined the distribution of writing processes over time, the following measures were extracted:

- **Number of events per subprocess and segment:** Each writing session was divided into three equal temporal intervals (T1, T2, T3), and the number of events in each subprocess was calculated per segment. This segmentation is common in writing process research (Roca de Larios et al., 2008; Garcés-Manzanera, 2021).

- **Proportion of events per segment:** The relative frequency of each subprocess across segments was computed to examine shifts in cognitive activity over time (Baaijen & Galbraith, 2018).
- **Event duration:** The length of each event, in milliseconds, was used descriptively to explore variation in processing time across subprocesses (Wengelin, 2006).

For **RQ2**, which focused on the relationship between pause location and writing subprocess, the dataset included:

- **Pause location:** Each pause was coded for its structural position relative to the emerging text, including categories such as within-word, between-word, between-sentence, and between-paragraph. These distinctions have been widely adopted in KLS-based research to reflect different levels of cognitive processing (Medimorec & Risko, 2017; Spelman Miller, 2000).
- **Pause–process association:** To examine the relationship between pause type and subprocess classification, a cross-tabulated matrix was created linking pause location to the type of process that followed. This analytical approach operationalises recent methodological proposals for context-sensitive pause analysis (Hall et al., 2022) and is consistent with empirical findings showing that pause location correlates with the type of cognitive activity that follows (Révész et al., 2023).
- **Proportional distribution:** The percentage of events preceded by each pause type was calculated for each subprocess, allowing for comparisons of how structural location influences cognitive transitions during writing.

For **RQ3**, which addressed pausing behavior across subprocesses, the following measures were computed:

- **Pauses per minute (ppm) per subprocess:** For each participant and subprocess, we divided the number of logged pauses by the minutes of active time spent in that subprocess. Rate-based pause indices are well established in keystroke logging research. Michel et al. (2020), for instance, calculated pause number per minute across different locations to capture fine-grained fluency patterns, and Vandermeulen et al. (2024) reported comparable per-minute rates across intervals. Barkaoui (2016) and Alexander (2019) also list the number of pauses per minute as a standard fluency measure. Medimorec and Risko (2017) adopt rate measures at text boundaries, linking increased pause rates to reduced fluency. While many studies apply thresholds (200 ms, 300 ms, 2 s) to separate motor from cognitive pauses (Medimorec & Risko, 2017; Van Waes & Leijten, 2015; Tian, Kim, & Crossley, 2024), the principle of normalizing pause frequency by time is widely used to gauge the density of writing activity.
- **Relative pause frequency per subprocess:** For each participant, we calculated the proportion of that participant's total pauses that occurred in each subprocess, yielding a within-writer distribution that sums to 1. This approach normalises raw pause counts and allows pausing to be interpreted as an allocation of effort. Proportional measures have precedent in pausing research: Gánem-Gutiérrez and Gilmore (2018) expressed episode frequency and duration as percentages of total time or episodes, precisely to avoid bias from individual variation in total output.

Michel et al. (2020) tracked indices proportionally across intervals, and Vandermeulen et al. (2024) reported distributions of pausing time in L1 and L2 writing. Earlier studies such as Manchón and Roca de Larios (2007) and Roca de Larios et al. (2008) likewise measured time spent on processes as percentages of composition time. These precedents confirm that relative pause frequency is a valid way to capture how subprocesses share pausing within writers.

- **Mean pause duration per subprocess:** For each participant and subprocess, we averaged the duration (ms) of pauses assigned to that subprocess. Mean pause duration is one of the most widely used temporal indicators of processing demands in writing. Spelman Miller (2000) showed that pause length increases with larger textual units, linking longer pauses to higher-level planning. Medimorec and Risko (2017) examined pause means across thresholds and boundaries, finding that longer pauses at sentences and paragraphs signalled more demanding processing, while shorter pauses within and between words reflected lexical retrieval and transcription. Révész et al. (2019) combined keystroke and eye-tracking evidence to show that mean pause duration by location mapped onto different subprocesses, with longer pauses tied to content planning and shorter pauses tied to linguistic encoding. Similar use of mean pause time as a proxy for cognitive effort is reported in Mohsen and Qassem (2020), Ke (2024), and Tian and Cushing (2025). As shown by evidence, these studies show that mean duration, when interpreted with pause location and context, may offer a robust indicator of the depth of processing.

All three measures were computed at the participant × subprocess level and then aggregated across participants for analysis. As a robustness check, the full set of analyses was repeated with a ≥ 2000 ms threshold to focus on long pauses only. Comparisons across subprocesses used Kruskal–Wallis tests with pairwise Wilcoxon tests and Holm adjustment (two-tailed $\alpha = .05$).

Data Analysis

Statistical analyses were conducted to address each of the three research questions using data extracted from the pause-informed classification system described above. All analyses were performed at the participant level, with subprocess-level aggregates used as input variables. As the dataset consisted of repeated measures across subprocesses and time intervals, non-parametric alternatives were used where normality assumptions were not met.

For **RQ1**, which examined the temporal distribution of writing subprocesses across the task, descriptive statistics were first computed for each subprocess within each time segment (T1, T2, T3). The proportion of events per subprocess was compared across segments using Friedman tests for repeated measures. Where significant differences emerged, Conover's post hoc comparisons were conducted with Holm-adjusted p-values to control the familywise error rate.

For **RQ2**, which investigated the association between pause location and writing subprocess, contingency tables were generated crossing pause types (within-word, between-word, between-sentence, between-paragraph) with subprocess categories. Pearson's chi-square tests of independence were used to examine whether the distribution of pause locations differed significantly across subprocesses. Standardized residuals were inspected to identify which cells contributed most to significant associations, and proportions within each subprocess category were reported to highlight the most frequent pause types.

For **RQ3**, which examined pausing behavior across subprocesses, we analyzed three measures: pauses per minute, mean pause duration, and relative pause frequency. Because distributions were non-normal, omnibus differences were tested with Kruskal–Wallis. When the omnibus test was significant, pairwise Wilcoxon rank-sum tests with Holm correction were applied. All tests were two-tailed with $\alpha = .05$. Effect sizes for Kruskal–Wallis are reported as rank ε^2 with 95% confidence intervals; for the ≥ 2000 ms sensitivity analysis we additionally report Kendall's W.

All analyses were conducted using JASP (Version 0.19.3.). Data were structured differently for each research question depending on the nature of the variables under investigation (e.g., temporal segmentation for RQ1; pause–process associations for RQ2). Effect sizes were reported where appropriate, including Kendall's W for Friedman tests and ε^2 for Kruskal–Wallis. We interpreted Kruskal–Wallis ε^2 using variance-based benchmarks of .01 (small), .06 (medium), and .14 (large), with .06–.13 treated as moderate, and we provide raw ε^2 and 95% CIs to aid interpretation. Given the absence of L2-specific ε^2 cutoffs, these values are presented descriptively.
Visual representations were used to support the statistical findings, and in the present paper we include heatmaps to illustrate co-occurrence patterns between pause types and processes.

Results

Temporal Distribution of Writing Processes (RQ1)

To examine how writing processes unfolded over time, a series of Friedman tests were conducted for each macroprocess (planning, formulation, and revision) and their corresponding subprocesses across three writing intervals. Table 2 provides the descriptive statistics for the proportion of total writing time devoted to each macroprocess across intervals, as well as the overall percentage of time allocated to each process throughout the task.

Table 2

Percentage of Total Writing Time Allocated to each Macroprocess and Distribution across Intervals

Macroprocess	Interval 1	Interval 2	Interval 3	Total (%)	SD (Across Intervals)
Planning	11.36	10.07	12.87	34.30	1.40
Formulation	17.39	17.58	17.85	52.82	0.24
Revision	4.37	4.71	3.80	12.88	0.46

Note. Total (%) represents the total percentage of time allocated to each macroprocess over the full writing task. SD indicates variability across the three intervals. N = 16.

For the macroprocesses, no statistically significant differences were observed across intervals: planning ($\chi^2(2) = 1.13$, $p = .57$, $W = .04$), formulation ($\chi^2(2) = 1.81$, $p = .40$, $W = .06$), and revision ($\chi^2(2) = 4.88$, $p = .09$, $W = .15$) - see Table 3. Although revision approached significance, post-hoc comparisons revealed no reliable pairwise differences after correction for multiple comparisons.

Table 3

Friedman Test with Conover Post Hoc Comparisons (Holm-adjusted) across Intervals for each Writing Process

Process	$\chi^2(2)$	p	Kendall's W
Planning			
– Global planning	0.88	.65	.03
– Local planning	4.13	.13	.14
Planning (total)	1.13	.57	.04
Formulation			
– Fluent formulation	1.63	.44	.05
– Lexical retrieval	1.63	.44	.05
– Problem-solving formulation	0.88	.65	.03
Formulation (total)	1.81	.40	.06
Revision			
– Subword revision	0.03	.98	~.00
– Word-level revision	1.63	.44	.05
– Above-word revision	0.80	.67	.04
Revision (total)	4.88	.09	.15

Note. Friedman tests computed across three intervals ($df = 2$, $N = 16$, except for Local Planning [N = 15] and Above-word Revision [N = 10]). Conover post hoc tests with Holm adjustment were used; no pairwise comparisons reached significance.

To further illustrate these patterns, Table 4 presents the proportion of time spent on each subprocess, showing its distribution across intervals and its contribution to the total time devoted to the task.

Table 4

Writing Time by Subprocess: Percentage Distribution across Intervals and Overall Task

Subprocess	Interval 1	Interval 2	Interval 3	Total (%)	SD (Across Intervals)
Global Planning	10.23	8.84	10.42	29.49	0.86
Local Planning	1.12	1.31	2.53	4.96	0.75
Fluent Formulation	11.82	11.39	11.43	34.64	0.24
Lexical Retrieval	2.46	2.38	2.61	7.45	0.12
Problem-solving Form.	3.12	3.81	3.81	10.74	0.40
Subword Revision	1.85	1.82	1.89	5.56	0.04
Word-level Revision	1.55	1.66	1.15	4.36	0.26
Above-word Revision	1.19	1.49	0.87	3.55	0.31

Note. Total (%) reflects the overall percentage of writing time devoted to each subprocess. SD indicates variability across intervals. N = 16, except Above-word Revision (N = 10).

None of the subprocesses - including fluent formulation, lexical retrieval, problem-solving formulation, global or local planning, or the three levels of revision - showed statistically significant changes across the writing intervals (all *ps* $> .10$). The effect sizes remained low throughout (Kendall's $W < .15$), suggesting that, on average, the relative proportion of each subprocess remained stable across the beginning, middle, and end of the writing task.

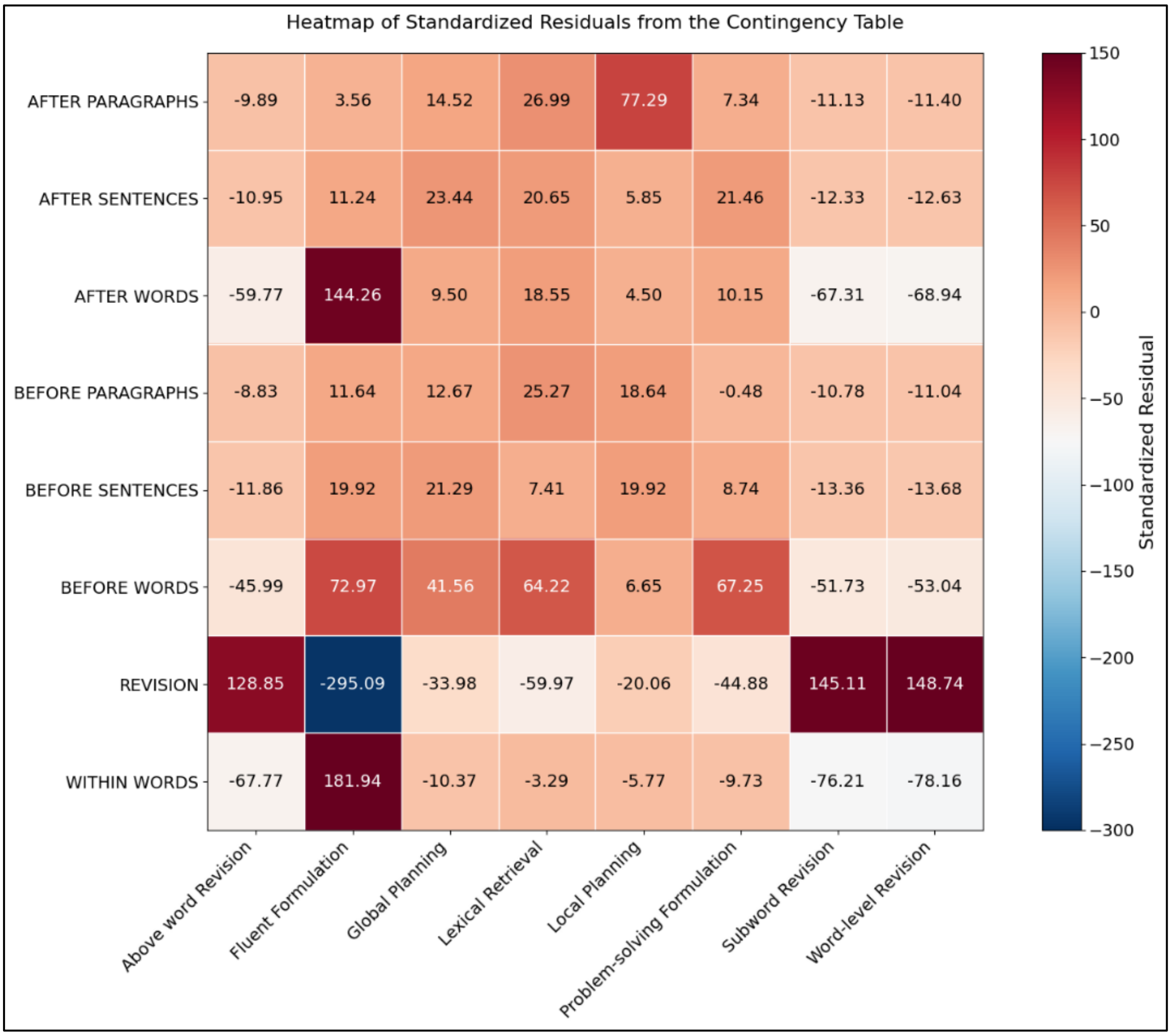
These findings suggest a relatively uniform temporal distribution of writing behaviors, indicating that adult L2 writers in this study did not substantially shift the allocation of planning, formulation, or revision efforts across different phases of text production.

RQ2

A chi-square test of independence revealed a significant association between pause location and writing subprocess, $\chi^2(49) = 126{,}118.72$, $p < .001$. The contingency coefficient indicated a strong relationship (C = .73), and Cramer's V suggested a medium effect size ($V = .41$). Given the magnitude of the chi-square value and the large sample size ($N = 108{,}309$), the distribution of pause locations differed systematically across subprocess categories.

Figure 1

Heatmap of Standardized Residuals from the Contingency Table Examining the Association between Pause Location and Writing Subprocess



Standardized residuals were inspected to identify the cells contributing most strongly to this association. These analyses were based on standardized residuals, where values $|z| > 1.96$ indicate significant deviations ($p < .05$). As shown in Figure 1, above-word and word-level revision pauses were almost exclusively located in the *Revision* subprocess ($|z| > 87$), whereas within-word pauses were disproportionately associated with *Fluent Formulation* ($z = 181.94$). High positive residuals were also observed for *Lexical Retrieval* and *Global Planning* in before-word positions ($z = 64.22$ and $z = 41.56$, respectively). Planning subprocesses also showed above-expected counts for between-unit pause locations, including before paragraphs and before sentences. These residuals reflect structural positions rather than any definitional overlap between location and subprocess.

Temporal Characteristics of Pauses across Writing Subprocesses (RQ3)

To examine differences in pausing behavior across subprocesses, we analyzed three measures: pauses per minute, mean pause duration, and relative pause frequency (the proportion of a participant's pauses occurring in each subprocess).

A Kruskal–Wallis test revealed significant differences in pauses per minute across subprocesses, $\chi^2(7) = 184.21$, $p < .001$, with a large effect size (rank $\varepsilon^2 = .41$, 95% CI [0.36, 0.48]). As shown in Table 5, pause rates were highest in formulation and revision subprocesses, particularly fluent formulation ($M = 162.23$ pauses/min), followed by above-word revision ($M = 119.87$ pauses/min) and word-level revision ($M = 96.93$ pauses/min). Planning processes displayed the lowest rates. Pairwise Wilcoxon tests with Holm adjustment confirmed that fluent formulation and above-word revision had significantly higher pause rates than all planning subprocesses, while lexical retrieval and global planning were significantly lower than all other subprocesses (Holm $ps < .001$).

A second Kruskal–Wallis test examined mean pause duration, which also differed significantly across subprocesses, $\chi^2(7) = 229.66$, $p < .001$, with a very large effect size (rank $\varepsilon^2 = .62$, 95% CI [0.57, 0.69]). Global planning showed the longest pauses on average ($M = 5272.74$ ms, $SD = 5873.78$), followed by problem-solving formulation ($M = 2001.63$ ms, $SD = 1567.50$). In contrast, fluent formulation and revision subprocesses were characterized by much shorter pauses (all < 650 ms). Post hoc comparisons confirmed that global planning and problem-solving formulation differed significantly from all formulation and revision subprocesses (Holm $ps < .001$).

Relative pause frequency also varied across subprocesses, $\chi^2(7) = 152.08$, $p < .001$, $W = .66$. On average, 44% of pauses occurred during fluent formulation, 24% during revision, and 20% during planning. Within planning, most pauses were concentrated in global planning, while within revision, above-word and word-level revisions dominated. This distribution indicates that although planning generated fewer pauses per minute, it still accounted for a large share of total pauses due to their long durations.

Table 5

Descriptive Statistics for Pauses per Minute, Mean Pause Duration, and Relative Pause Frequency by Subprocess

Subprocess	N	Pauses/min (M)	SD	Pause mean duration (M)	SD	Rel. pause frequency (M)	SD
Fluent formulation	16	162.23	18.42	270.57	28.96	0.44	0.08
Above-word revision	16	119.87	56.77	558.75	295.44	0.14	0.05
Word-level revision	16	96.93	33.72	631.54	294.58	0.07	0.03
Subword revision	16	63.73	18.06	925.86	272.41	0.03	0.02
Lexical retrieval	16	39.23	2.30	1386.35	43.38	0.06	0.02
Problem-solving formulation	16	7.33	7.43	2001.63	1567.50	0.08	0.03
Local planning	16	4.30	4.54	802.26	800.76	0.04	0.02
Global planning	16	1.46	1.93	5272.74	5873.78	0.14	0.06

Note. Rate = number of inter-event delays per minute of active time in the subprocess. Mean inter-event delay expressed in milliseconds. N = number of subprocess observations. Pairwise comparisons use Wilcoxon's test with Holm correction; significance refers to Holm-adjusted p-values.

Sensitivity check (≥ 2000 ms)

As a robustness analysis, we repeated the tests with a 2000 ms threshold to restrict the analysis to long pauses (see Table 6). The omnibus test for pauses per minute remained significant, $\chi^2(7) = 80.63$, $p < .001$, $W = .72$. Under this filter, mean durations were highest for global planning (M = 10,472.49 ms) and problem-solving formulation (M = 4096.05 ms). Pause rates for fluent formulation fell to zero, and those for revision subprocesses were close to zero. Relative pause frequency shifted toward planning and problem-solving processes, with global planning (M = .37) and problem-solving formulation (M = .40) accounting for most of the observed long pauses. In contrast, formulation and revision subprocesses showed only minimal contributions once the threshold was applied.

Table 6

Descriptive Statistics for Pauses per Minute, Mean Pause Duration, and Relative Pause Frequency by Subprocess (≥ 2000 ms Threshold)

Subprocess	Pauses/min M	SD	Mean duration (ms) M	SD	Relative freq M	SD
Fluent formulation	~0.00	0.00	—	—	~0.00	0.00
Above-word revision	0.03	0.02	2300.67	1200.32	0.02	0.01
Word-level revision	0.05	0.03	2467.14	1183.59	0.03	0.02
Subword revision	0.01	0.00	2250.46	1121.24	0.01	0.00
Lexical retrieval	0.04	0.03	3267.89	1510.42	0.07	0.02
Problem-solving formulation	0.13	0.06	4096.05	1664.17	0.40	0.08
Local planning	0.06	0.04	2746.82	1423.55	0.10	0.04
Global planning	0.15	0.07	10472.49	5021.77	0.37	0.09

Discussion

Temporal Distribution of Writing Processes (RQ1)

The analyses conducted for RQ1 revealed a stable temporal distribution of writing processes across the three intervals. No statistically significant differences were observed in the proportion of time devoted to planning, formulation, or revision, and the same pattern applied to the corresponding subprocesses. Although the omnibus test for revision approached significance (p = .09), pairwise comparisons yielded no effects that remained significant after correction. Studies on source-based writing report that recursivity peaks in the middle rather than at the start or end, which helps interpret the flat temporal profile observed here (Tarchi et al., 2024).

This absence of temporal shifts contrasts with findings from studies reporting differentiated allocation of cognitive effort over time, particularly in L1 contexts or among experienced L2 writers (e.g., Xu & Xia, 2021; Breetvelt et al., 1994). However, other research with younger or less proficient L2 writers has also documented similar uniformity, often linked to limited procedural knowledge or underdeveloped strategic control (López et al., 2019; Olive et al., 2009). In such cases, writers tend to sustain the same operational mode throughout the task, without allocating distinct phases for planning or revision.

This trend appeared consistently across both broader processes and more specific subprocesses, with planning standing out in particular. Although theoretical models have often assumed that global planning occurs early in the writing task (Hayes, 1996; Kellogg, 1996; Révész et al., 2023), the results here do not support that expectation. Large-sample keystroke evidence shows that timing, higher and lower order pauses, production speed, and revision vary by language and genre, so an early planning peak is not universal (Vandermeulen et al., 2024). There was no clear concentration of planning activity at the beginning. Other studies involving keystroke data from adolescent and adult L2 writers have described similar distributions, where markers of planning are scattered throughout the task rather than grouped in the initial phase (Vandermeulen et al., 2024; Tillema, 2012).

Dividing the writing task into three equal parts is a strategy often used in keystroke studies to support comparisons across datasets (e.g., Xu, 2018; Tarchi et al., 2023; Vandermeulen et al., 2024). In this case, the same procedure was followed. However, this type of segmentation does not reflect how content is structured or how cognitive demands may shift throughout the task. Although it offers a workable solution, any conclusions drawn from temporal distribution should be treated with care, especially given the possibility that writing processes unfold in a recursive or non-sequential manner (Galbraith, 2009; Conijn et al., 2024, 2025).

When considered as a whole, the findings suggest that the adult L2 writers in this study did not show major changes in how they distributed cognitive effort over time. The levels of activity linked to planning, formulation and revision remained relatively steady across the initial, middle and final parts of the task.

Pause Location and Writing Subprocesses

The results obtained for RQ2 support the interpretation that pause location is meaningfully related to the subprocess active during the writing event. Rather than reflecting incidental temporal gaps, pause positions corresponded consistently with distinct cognitive operations. This reinforces previous work suggesting that pause data, when interpreted with contextual information, can serve as an indicator of process sequencing during composition (Medimorec & Risko, 2017; Hall et al., 2024). Within-word and before-word pauses occurred primarily during formulation, especially in fluent formulation and lexical retrieval. This distribution replicates findings from L2 writing studies, where low-level disruptions often stem from lexical access difficulties and syntactic encoding demands (Roca de Larios et al., 2006; Révész et al., 2019). These outcomes are in line with cognitive models that locate formulation at the interface between linguistic retrieval and idea generation (Flower & Hayes, 1981; Hayes, 2012). The classification system applied here, based on the spatial and functional context of events rather than surface fluency, proved sensitive to such associations.

Global planning has frequently been associated with pauses occurring before sentences and paragraphs, positions that have long been linked to conceptual structuring and discourse-level organization (Wengelin, 2006; Wengelin et al., 2009). This interpretation is consistent with Kellogg's (1996) model, which highlights the particularly high cognitive demands of conceptual planning compared with other writing operations. Research has also indicated that more advanced writers are likely to pause at structurally salient locations when preparing to initiate new macro-units (Révész et al., 2023). Although pause duration was not examined here, the positions observed correspond to those widely regarded as cognitively demanding in writing research.

Pauses classified as revision were most often located next to edit operations such as deletions, insertions, or cursor repositioning. Since the classification system defines revision through these very operations, this pattern serves as an internal validation of the method rather than an independent observation. Word-level and above-word revision pauses were rarely situated at the point of inscription where ongoing text production unfolds. This spatial separation reflects a shift in attention away from formulation and supports the distinction between precontextual revisions at the leading edge and contextual revisions within the produced text (Lindgren & Sullivan, 2006). It also aligns with findings that higher-level revisions typically target completed text segments (Révész et al., 2019). While some revisions do occur during formulation at the inscription point, especially at the subword level when correcting typographic errors, pauses linked to word- or clause-level revision are usually detached from the flow of fluent text production.

Together, these patterns are consistent with process models that define writing as a recursive and context-sensitive activity, rather than a linear sequence of stages (Hayes, 2012; Galbraith, 2009). The classification system employed here, grounded in pause location and local event structure, proved sensitive to the functional segmentation of writing behavior. Alongside conventional thresholded approaches, it captured meaningful associations between cognitive subprocesses and their temporal-spatial signatures, offering a principled method for identifying writing processes in L2 contexts.

Temporal Characteristics of Writing Subprocesses

RQ3 asked whether subprocesses can be differentiated temporally as well as structurally. Group differences in pauses per minute and mean pause duration, together with relative pause frequency within writers, suggest that cognitive demands are uneven across subprocesses during L2 composing (Hayes, 2012; Kellogg, 1996). Because pauses per minute count all logged pauses without a lower bound, high rates in revision reflect dense sequences of short production-level events rather than many long pauses.

A contrast was visible between planning and revision. Global planning involved long pauses that occurred only occasionally, consistent with evidence that pauses at higher textual units, such as before sentences and paragraphs, often index conceptually demanding operations where ideas are generated and organised prior to formulation (Medimorec & Risko, 2017; Révész et al., 2019; Manchón & Roca de Larios, 2007). These pauses have typically been described as pre-verbal processing and are often concentrated at the start of composing (Ong, 2014; Roca de Larios et al., 2008; López et al., 2019), although no early peak appeared in the present data. In contrast, revision at the subword, word, and above-word levels was characterised by frequent but shorter pauses, a profile that reflects local corrections such as spelling, grammar, or single-word substitutions often executed in brief bursts and accompanied by cursor moves or overwriting (Barkaoui, 2016; Xu, 2018; Conijn et al., 2024). While meaning-related revisions can demand more sustained engagement, most local edits in L2 writing tend to be quick and recurrent (Révész et al., 2019; Mohsen & Qassem, 2020). This divergence between long, infrequent planning pauses and short, dense revision pauses is consistent with research linking pause location and timing to different levels of processing, with higher-order conceptual work associated with longer pauses at larger textual boundaries and lower-order corrections linked to shorter pauses at smaller units (Spelman Miller, 2000; Medimorec & Risko, 2017).

Problem-solving formulation occupied an intermediate position. Pauses in this subprocess were longer on average than those in fluent formulation but shorter than those observed in extended planning, with mean durations close to two seconds in the present data. Their frequency was lower than in formulation and revision, yet they accounted for about 8% of all pauses. This profile is consistent with prior work showing that moments of re-analysis or restructuring during writing impose raised cognitive demands (Roca de Larios et al., 2006; Révész et al., 2023). It is also compatible with dual-process accounts of writing in which spontaneous text generation and more controlled operations interact dynamically during formulation (Galbraith, 2009).

Within formulation, a further contrast emerged between fluent transcription and lexical retrieval. Fluent formulation was characterised by short pauses and dense sequences of text production, consistent with rapid idea-to-text translation and uninterrupted bursts of transcription (Hayes, 2012; Kim, 2022). In the present data, this subprocess showed the highest pause rate per minute, although the pauses themselves were brief, which indicates that the elevated frequency reflects the density of production-level activity rather than prolonged cognitive interruptions. Lexical retrieval involved longer pauses than fluent formulation, consistent with the added effort of accessing and selecting lexis. In the present dataset, pause rates for lexical retrieval did not exceed those for fluent formulation, and the pairwise comparison on rates was not significant after correction, yet the difference in average duration pointed toward greater cognitive load. Prior keystroke studies similarly associate pauses at word boundaries with lexical encoding demands, while longer pauses at higher boundaries are more closely linked to discourse planning (Medimorec & Risko, 2017; Révész et al., 2019; Mohsen & Qassem, 2020). Taken together, these distinctions reinforce the need to treat formulation not as a single process but as a continuum ranging from fluent transcription to effortful retrieval and restructuring.

The relative pause frequency analysis situates these patterns within writers. When all pauses were considered, most occurred in fluent formulation and revision, with smaller proportions in planning. This reflects the dominance of short, frequent pauses tied to transcription, lexical encoding, and local corrections during active text production (Medimorec & Risko, 2017; Barkaoui, 2016; Xu, 2018). Once the analysis was restricted to pauses of 2000 ms or more, the distribution shifted toward planning-related activity: long pauses clustered in global planning and problem-solving formulation, while fluent formulation fell to zero and revision contributed very little. In combination, the unfiltered profile captures the density of production-level activity, whereas the filtered profile highlights the less frequent but conceptually oriented processing associated with higher-level planning and restructuring (Hall et al., 2024; Révész et al., 2019; Tian & Cushing, 2025).

Consistency across the three measures strengthens confidence in the classification. Rates, durations, and within-writer shares pointed in the same direction, and the ≥ 2000 ms sensitivity check reproduced the shift toward planning-related activity. Pauses were interpreted in context rather than in isolation, taking into account their textual location and the actions immediately before and after the pause, which is essential because pause functions vary with boundary and surrounding behavior (Lindgren & Sullivan, 2006; Medimorec & Risko, 2017; Révész, Michel, & Lee, 2019; Hall et al., 2024). The recurrence of similar temporal patterns across writers suggests systematic shifts of attention during composing, consistent with prior work on process regularities in L2 writing (Barkaoui, 2019; Xu & Qi, 2017). At the same time, caution is warranted: pauses are multiply determined and any single metric is an indirect indicator. Read in context, however, time-based measures provide a clearer window on how writing activity unfolds and where cognitive effort is concentrated (Hall et al., 2024; Tian & Cushing, 2025).

Potential pedagogical uses follow from these distributions. Pauses within or immediately before words may signal lexical access strain and point to instruction targeting vocabulary growth, morphological control, or spelling. Sparse evidence of long planning pauses in sections that call for global structuring may indicate limited use of higher-order organisation strategies. Process logging can surface these patterns for formative feedback and targeted support, for example through process reports and dashboards that visualise pausing and revision profiles for reflection and coaching (Ranalli et al., 2018; Vandermeulen et al., 2024; Tian & Cushing, 2025).

Conclusion

This study examined how adult L2 writers distributed their cognitive effort across a timed writing task, drawing on keystroke logging data and a pause-informed classification system that distinguished among planning, formulation, and revision, as well as their constituent subprocesses. The system, which combined temporal thresholds, structural pause location, and local event context, enabled the automated identification of writing behavior grounded in established cognitive models (Hayes, 2012; Galbraith, 2009).

The results indicated that planning, formulation, and revision did not redistribute systematically across the three task intervals. Rather than an initial planning stage followed by end-loaded revision, as assumed in earlier linear models (Breetvelt et al., 1994), traces of planning and reworking appeared throughout the composing process, even under time pressure. This recursive profile echoes accounts of writing as dynamic and overlapping (Flower & Hayes, 1981; Sommers, 1978; Roca de Larios et al., 2008; Vandermeulen et al., 2024). Pause characteristics added further differentiation across subprocesses. Rates were highest in fluent formulation and local revision, reflecting the density of production-level activity, whereas durations were longest in global planning and problem-solving formulation, consistent with the greater effort required for conceptual structuring and controlled reformulation (Medimorec & Risko, 2017; Roca de Larios et al., 2006; Révész et al., 2019). Relative frequency helped situate these measures within writers. When all pauses were included, the largest shares clustered in formulation and revision, but when only pauses longer than two seconds were considered, the distribution shifted toward global planning and problem-solving formulation. This divergence illustrates how unfiltered measures capture the micro-pauses of continuous production, while filtered measures isolate conceptually oriented activity. This convergence across rate, duration, and relative frequency supports the robustness of the classification system and shows that context-sensitive interpretation of pausing can offer a clearer view of underlying processes (Spelman Miller, 2000; Medimorec & Risko, 2017; Hall et al., 2024).

Methodologically, the study shows that a classification system based on temporospatial criteria and local event context can capture recursive L2 writing at scale without relying on linear models or post-hoc coding (Leijten & Van Waes, 2013; Spelman Miller, 2000). The procedure adopted in this study offers a principled way of interpreting keystroke data while retaining the dynamic nature of composing. At the same time, important limitations need to be acknowledged. The sample was modest (N = 16), the processes inferred from the logs are indirect, and there was no concurrent verbal report or validation against expert judgment. Although a small-scale manual check was conducted, further work is needed to strengthen interpretive claims. Future studies should test the classification across larger and more varied samples, and triangulate keystroke data with complementary sources such as think-aloud protocols, stimulated recall, or expert ratings to refine the mapping between logged behavior and underlying cognitive activity (cf. Hall et al., 2024).

Finally, the findings point to pedagogical applications beyond research. Keystroke logging can render visible aspects of composing that are typically inaccessible, such as disproportionate effort devoted to lexical retrieval or the absence of sustained planning. Making such patterns explicit creates opportunities for formative feedback and the development of metacognitive strategies (Ranalli et al., 2018; Vandermeulen et al., 2024). With further refinement, process dashboards could present these data in ways that assist teachers in diagnosing difficulties and adapting instruction, while enabling learners to reflect systematically on their own composing practices. In this respect, keystroke logging has potential not only as an analytic instrument but also as a pedagogical resource for supporting L2 writing development.

ORCID

https://orcid.org/0000-0002-1789-9046

Acknowledgements

I am grateful to Dr Rosa Manchón, Dr Lourdes Cerezo, and Dr Raquel Criado for granting permission to reuse data collected within the AEI-funded research project (reference: PID2022-137544NB-I00).

Funding

This study was done under the auspices of one competitive program of research: I+D+i Grants PID2022-137544NB-I00, funded by the Spanish National Research Agency and the Ministry of Science, Innovation and Universities.

Ethics Declarations

Competing Interests

The author declares no competing interests.

Rights and Permissions

Open Access

References

Abdel Latif, M. M. (2008). A state-of-the-art review of the real-time computer-aided study of the writing process. *International Journal of English Studies, 8*(1), 29–50. Retrieved from https://revistas.um.es/ijes/article/view/49081

Baaijen, V. M., & Galbraith, D. (2018). Discovery through writing: Relationships with writing processes and text quality. *Cognition and Instruction, 36*(3), 1–25. https://doi.org/10.1080/07370008.2018.1456431

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. https://doi.org/10.1111/modl.12316

Barkaoui, K. (2019). What keystroke logging can reveal about second language writing processes. In E. Lindgren & K. P. H. Sullivan (Eds.), *Observing writing* (pp. 118–139). Brill.

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

Chenoweth, N. A., & Hayes, J. R. (2001). Fluency in writing: Generating text in L1 and L2. *Written Communication, 18*(1), 80–98. https://doi.org/10.1177/0741088301018001004

Conijn, R., Rossetti, A., Vandermeulen, N., & Van Waes, L. (2025, preprint). *Phase to phase: Towards an automated procedure to identify phases in writing processes using keystroke data*. SSRN. https://doi.org/10.2139/ssrn.4993558

Conijn, R., Dux Speltz, E., & Chukharev-Hudilainen, E. (2024). Automated extraction of revision events from keystroke data. *Reading and Writing, 37(2),* 483–508. https://doi.org/10.1007/s11145-021-10222-w

Conijn, R., Dux Speltz, E., van Zaanen, M., Van Waes, L., & Chukharev-Hudilainen, E. (2022). A product- and process-oriented tagset for revisions in writing. *Written Communication, 39*(1), 97–128. https://doi.org/10.1177/07410883211052104

Flower, L., & Hayes, J. R. (1981). A Cognitive Process Theory of Writing. *College Composition and Communication, 32(4),* 365–387. https://doi.org/10.2307/356600

Galbraith, D., & Baaijen, V. M. (2019). Aligning keystrokes with cognitive processes in writing. In E. Lindgren & K. Sullivan (Eds.), *Observing Writing (Vol. 38, pp. 306–325)*. Brill. https:// doi. org/ 10. 1163/ 97890 04392 526_ 015

Galbraith, D. (2009). Cognitive models of writing. In R. Beard, D. Myhill, M. Nystrand, & J. Riley (Eds.), *The SAGE handbook of writing development* (pp. 161–178). SAGE.

Garcés-Manzanera, A. (2021). An exploratory study of primary school children's writing processes in digital environments: The use of models as written corrective feedback [Doctoral dissertation, University of Murcia]. DIGITUM repository. http://hdl.handle.net/10201/114544

Garcés-Manzanera, A. (2025, forthcoming). *Digital L2 writing and keystroke logging software: Observing writing behaviors*. Nova Science Publishers.

Hall, S., Baaijen, V. M., & Galbraith, D. (2024). Constructing theoretically informed measures of pause duration in experimentally manipulated writing. *Reading and Writing, 37*(2), 329–357. https://doi.org/10.1007/s11145-022-10284-4

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

Hayes, J. R. (2012). Modeling and remodeling writing. *Written Communication*, *29*(3), 369–388. https://doi.org/10.1177/0741088312451260

Ke, Y. (2024). Examining simultaneous pausing on the cognitive writing process: A micro-formative writing assessment. *Current Psychology, 43*(39–50). https://doi.org/10.1007/s12144-023-04429-z

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.

Kim, Y.-S. G. (2022). Do written language bursts mediate the relations of language, cognitive, and transcription skills to writing quality? *Written Communication, 39*(2), 200–227. https://doi.org/10.1177/07410883211068753

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

Lindgren, E., & Sullivan, K. P. H. (2006). Writing and the analysis of revision: An overview. In K. P. H. Sullivan & E. Lindgren (Eds.), *Computer key-stroke logging and writing* (pp. 31–44). Brill. https://doi.org/10.1163/9780080460932_004

López, P., Torrance, M., & Fidalgo, R. (2019). The online management of writing processes and their contribution to text quality in upper-primary students. *Psicothema*, *31*(3), 311–318. https://doi.org/10.7334/psicothema2018.326

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

Manchón, R. M. (2011). Situating the learning-to-write and writing-to-learn dimensions of L2 writing. In R. M. Manchón (Ed.), *Learning-to-write and writing-to-learn in an additional language* (pp. 3–14). John Benjamins.

Medimorec, S., & Risko, E. F. (2017). Pauses in written composition: On the importance of where writers pause. *Reading and Writing*, *30*, 1267–1285. https://doi.org/10.1007/s11145-017-9723-7

Michel, M., Révész, A., Lu, X., Kourtali, N. E., 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

Mohsen, M. A., & Qassem, M. (2020). Analyses of L2 Learners' Text Writing Strategy: Process-Oriented Perspective. *Journal of Psycholinguistic Research*, *49*, 435–451.

Olive, T., Alves, R. A., & Castro, S. L. (2009). Cognitive processes in writing during pause and execution periods. *European Journal of Cognitive Psychology*, *21*(5), 758-785. https://doi.org/10.1080/09541440802079850

Ong, J. (2014). How do planning time and task conditions affect metacognitive processes of L2 writers? *Journal of Second Language Writing*, *23*, 17–30. https://doi.org/10.1016/j.jslw.2013.10.002

Ranalli, J., Feng, H. H., & Chukharev-Hudilainen, E. (2018). Exploring the potential of process-tracing technologies to support assessment for learning of L2 writing. *Assessing Writing*, *36*, 77–89. https://doi.org/10.1016/j.asw.2018.03.007

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*(3), 680–709. https://doi.org/10.1017/S0272263123000074

Roca de Larios, J., Manchón, R. M., Murphy, L., & Marín, J. (2006). Generating text in native and foreign language writing: A temporal analysis of problem-solving formulation processes. *The Modern Language Journal, 90*(1), 100–114. https://doi.org/10.1111/j.1540-4781.2006.00387.x

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*(1), 30–47. https://doi.org/10.1016/j.jslw.2007.08.005

Sasaki, M., Mizumoto, A., & Murakami, A. (2018). Developmental trajectories in L2 writing strategy use: A self-regulation perspective. *The Modern Language Journal, 102*(2), 292–309. https://doi.org/10.1111/modl.12469

Spelman Miller, K. (2000). Academic writers on-line: Investigating pausing in the production of text. *Language Teaching Research*, *4*(2), 123–148. https://doi.org/10.1177/136216880000400203

Sommers, N. I. (1978). *Revision in the composing process: A case study of college freshman and experienced adult writers* (Unpublished doctoral dissertation). Boston University.

Tarchi, C., Villalón, R., Vandermeulen, N., Casado-Ledesma, L., & Fallaci, A. P. (2023). Recursivity in source-based writing: A process analysis. *Reading and Writing*, *37*, 2571–2593. https://doi.org/10.1007/s11145-023-10482-8

Tian, Y., & Cushing, S. T. (2025). Exploring the application of keystroke logging techniques to research in second language (L2) writing. *Research Methods in Applied Linguistics, 4*(1), Article 100179. https://doi.org/10.1016/j.rmal.2024.100179

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

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

Vandermeulen, N., Lindgren, E., Waldmann, C., & Levlin, M. (2024). Getting a grip on the writing process: (Effective) approaches to write argumentative and narrative texts in L1 and L2. *Journal of Second Language Writing*, *65*, Article 101113. https://doi.org/10.1016/j.jslw.2024.101113

Wengelin, Å., Torrance, M., Holmqvist, K., Simpson, S., Galbraith, D., Johansson, V., & Holmqvist, K. (2009). Combined eyetracking and keystroke-logging methods for studying cognitive processes in text production. *Behavior Research Methods, 41(2),* 337–351. https://doi.org/10.3758/BRM.41.2.337

Wengelin, Å. (2006). Examining pauses in writing: Theories, methods and empirical data. In K.P.H. Sullivan & E. Lindgren (Eds.), *Computer Key-Stroke Logging and Writing: Methods and Applications* (pp. 107-130). Elsevier. https://doi.org/10.1163/9780080460932_008

Xu, C. (2018). Understanding online revisions in L2 writing: A computer keystroke-log perspective. *System*, *78*, 104–114. https://doi.org/10.1016/j.system.2018.08.007

Xu, C., & Ding, Y. (2014). An exploratory study of pauses in computer-assisted EFL writing. *Language Learning & Technology*, *18*(3), 80–96. https://doi.org/10.64152/10125/44385

Xu, C., & Qi, Y. (2017). Analyzing pauses in computer-assisted EFL writing: A computer-keystroke-log perspective. *Journal of Educational Technology and Society*, *20*, 24–34. https://www.jstor.org/stable/26229202

Xu, C., & Xia, J. (2021). Scaffolding process knowledge in L2 writing development: Insights from computer keystroke log and process graph. *Computer Assisted Language Learning*, *34*(4), 583–608. https://doi.org/10.1080/09588221.2019.1632901