



Methods for studying the writing time-course

Mark Torrance¹ · Rianne Conijn²

Accepted: 17 October 2023
© Crown 2023

Abstract

The understanding of the cognitive processes that underlie written composition requires analysis of moment-by-moment fluctuation in the rate of output that go beyond traditional approaches to writing time-course analysis based on, for example, counting pauses. This special issue includes 10 papers that provide important new tools and methods for extracting and analyzing writing timecourse data that go beyond traditional approaches. The papers in this special issue divide into three groups: papers that describe methods for capturing and coding writing timecourse data from writers producing text either by hand or by keyboard, papers that describe new statistical approaches to describing and drawing inferences from these data, and papers that focus on analysis of how a text develops over time as the writer makes changes to what they have already written.

Keywords Writing processes · Text production · Keystroke logging · Handwriting · Text-analysis · Cascading

This special issue presents papers that describe new or underused methods for capturing and analysing writing as an activity that proceeds in real time. There are broadly three overlapping reasons why as researchers we might want to do this: (1) because

Work on this special issue was supported, in part, by a European Association for Research on Learning and Instruction Emerging Fields Group grant to the first author. Neither author has conflict of interest to report.

✉ Mark Torrance
mark.torrance@ntu.ac.uk
Rianne Conijn
m.a.conijn@tue.nl

¹ Nottingham Trent University, Nottingham, UK

² Eindhoven University of Technology, Eindhoven, Netherlands

we are interested in how writers strategically control their own engagement in the various activities necessary to generate ideas and structure them on the page, (2) because we want to understand the psycholinguistic mechanisms that underlie the moment-by-moment processing that creates written language, or (3) because we want to understand how the text itself develops and mutates over time. In the first half of this introduction to the special issue we will briefly discuss each of these reasons and associated methodological challenges. In the second half we provide an overview of the 10 papers that follow.

Written products do not emerge fully formed from writers' pens or keyboards. What the reader finally sees has developed over time. Writers start with just a goal. This may be very specific ("write a 150 word abstract for my paper") or very general ("write a story that will please my teacher"). Either way, both content and expression must then be generated in real time as writing progresses. The language used in the final text – syntax and word-choice – and, to varying degrees, the message that it communicates are not represented in the writer's mind when they start to write. They are emergent outputs of a real-time process.

A number of researchers, dating back at least to Emig (1971), have taken as their starting point the assumption that how a writer organizes different writing activities (or "subprocesses") affects the quality of what they produce (e.g., Braaksma et al., 2004; Breetvelt et al., 1994; Flower & Hayes, 1980; Hayes & Flower, 1980b, 1980a; Van Den Bergh & Rijlaarsdam, 2001). Research in this tradition has three defining characteristics. First theory about writing processes is inferred from what writers said when they were asked to think aloud while writing. This is an approach to data collection that has its roots in very early psychological research but was revived and given respectability by researchers studying the processes by which people solve problems (Ericsson & Simon, 1984; Newell & Simon, 1972). Following from this, a second characteristic of this understanding of writing process is a focus on higher-level thinking and reasoning activity – the kind of activity that might be exposed in writers' think-aloud protocols. This is explicit in Flower and Hayes' (1980) problem-solving model of writing, and implicit in subsequent work. This leads to a third assumption, namely that writers, or at least skilled writers, deliberately and explicitly orchestrate their writing processes: When a writer stops to plan what to say next, or to review and make changes to what they have already written, is under their executive control. From an educational point of view, this position is attractive. If it is the case that writers are able to decide when to engage in specific writing subprocesses, then these decisions are available to manipulation through instruction: Writers can be taught explicit strategies that change what they do during composition in ways that benefit the quality of their text (e.g., Graham et al., 2005).

However, as an approach to understanding what happens in a writer's mind as they compose text, an orchestration-of-subprocesses account is incomplete for the simple reason that, as with spoken language production, much of the processing associated with producing text occurs rapidly and implicitly. Newell (1992) made a useful distinction between mental activity that occurs within the *Intendedly Rational Band*, occurring at timescales above 10 s, and activity that occurs within the *Cognitive Band* with timescales around 1 s. Research that seeks to understand the writing process as an orchestrated problem-solving activity focuses explanation on activity within the

Intendedly Rational Band. The focus of explanation is on how retrieved information is used to achieve specific, complex goals rather than the moment-by-moment processing by which it was retrieved in the first place. Cognitive band operations, on the other hand, although also goal directed, are implicit and outside of our control. Consider for example the processing necessary for writing the name of an everyday object, or retrieving the spelling and syntax for writing a sentence.

Our present purpose is not to discuss how essential this higher-level, intendedly-rational processing is to successful text production (but see Torrance, 2016). However we note that, at least for reasonably competent writers, composition often occurs remarkably fluently, with very few hesitations of a duration that would be consistent with intendedly-rational processing. In a reanalysis of keystroke data from adolescent writers writing short argumentative essays (Rønneberg et al., 2022) using methods described by Roeser et al. (2021), Roeser and Torrance (in preparation) found that writers rarely hesitated at sentence boundaries, with over 50% of sentences preceded by very short pauses (mean around 430 ms) and with the mean of the remainder of pauses at only around 1.2 s. These and similar findings point towards much of the mental activity associated with composition, including the relatively complex processing required to plan sentences, occurring as a result of a cascade of processes that occur partly in parallel and largely without executive control (Olive, 2014; van Galen, 1991). When writers move from one sentence to another without hesitation this is due to planning the next sentence, to some extent, while completing output of the previous sentence.

The second reason why we might want to explore the writing timecourse is therefore because in contexts where mental activity is not available to introspection, theories about process can be tested by measuring how long it takes people to perform particular cognitive tasks. This results in an approach to understanding writing timecourse data that is quite different from the research focused on writing as an orchestrated problem-solving activity that we have been discussing. Researchers interested in the orchestration of writing processes focus on what happens when; in the sequencing of different composing activities and the proportion of total time for which they are engaged. Researchers whose interest is in the fundamental cognitive processes that make text composition possible are more concerned with moment-by-moment fluctuation in rate of output: The duration of the hesitation before specific output (starting to write a word, for example) is, with an important qualification that we return to below, a measure of the complexity or difficulty of the mental operations that make that output possible.

This use of time data is a mainstay of cognitive-experimental language research. For example there is a long history of research in spoken production in which researchers ask participants to produce words or sentences in response to picture stimuli and measure response latency – the time from stimulus presentation to utterance onset (for early examples see Levelt & Maasen, 1981; Oldfield & Wingfield, 1964). More recently a similar experimental literature has emerged in written production (e.g., Bonin & Fayol, 2000; Bonin et al., 2012; Pinet & Nozari, 2018; Roeser et al., 2019). The ubiquity, outside of primary school, of typewritten production and the easy availability of software that records the timing of each keypress during composition (principally *InputLog*; Leijten & Van Waes, 2013) has led to a rapidly expand-

ing composition timecourse literature with a main focus not on the orchestration of subprocesses but on when and where writers hesitate during production. In the three years up to 2022 the Social Science Citation Index reports 40 journal papers that describe research exploring composition processes using keystroke logging methods, compared to 26 in 2017 to 2019, and 7 in 2014 to 2016.

Nearly all of these papers and, with one exception, the papers in this special issue focus on *spontaneous* text production tasks (Gernsbacher & Givón, 1995). These are tasks – essay or narrative writing for example – in which participants write multiple sentences in response to a topic statement. Analysing and interpreting chronometric data in this context – latencies before sentences and words, for example – poses several substantial challenges that are not faced by researchers who analyse response latencies in picture naming experiments.

First, and most obviously, spontaneous composition does not allow the kind of experimental control that is available in word or sentence production tasks. The latency before, for example, a mid-sentence word is potentially influenced by multiple factors including the length, frequency, and regularity of the word that is about to be produced. In an experimental context these factors can be crossed or controlled. This is not possible when the writer decides what to write. Keystroke logging or handwriting studies of writers producing essays or narratives are, therefore, essentially observational.

This raises a second issue. Because the researcher is observing rather than controlling what text is produced, they then require analytic methods for identify the linguistic units within the text over which planning might be hypothesized to scope. Words are obvious candidate planning units: It is reasonable to assume that the interkey interval between pressing the spacebar and pressing the initial key of a word is determined – sometimes and in part – by the difficulty or complexity of the processing that is necessary to mentally prepare the word. Orthographic sentence boundaries – sentence boundaries that are marked by sentence-terminating punctuation – are similarly easy to identify. Latencies at the start of a sentence are likely, again sometimes and in part, to be determined by the extent and complexity of the syntactic planning necessary prior to starting a new main clause (but see Roeser et al., 2019). Automatically identifying keystrokes that occur within words, before mid-sentence words, and at the start of sentences is relatively straightforward, and a number of studies have compared keystroke latencies at these locations (e.g., Conijn et al., 2019; Medimorec & Risko, 2017; Mohsen, 2021; Torrance et al., 2016; Wengelin, 2002). Fewer studies have gone beyond this distinction to explore planning of linguistically-defined text spans (e.g., T-units, finite and non-finite clauses; but see Ailhaud & Chenu, 2017; Ailhaud et al., 2016; Chukharev-Hudilainen et al., 2019; Leijten et al., 2019).

A third question with which researchers studying moment-by-moment writing activity must grapple is how best to describe and model the latency data that they are collecting. Understanding this is fundamental to drawing inferences about underlying process. The issue here is that distribution of inter-keystroke intervals, and of similar measures taken from handwriting, is strongly positively skewed. Therefore traditional measures of central tendency and dispersion – means and standard deviations, or medians and ranges – are misleading. All response time data, including experimental data, are positively skewed simply because there are tight limits on how

quick a participant's response can be. These are imposed by the cognitive system but also simply by the fact that response times cannot be less than zero. There are, however, no corresponding upper limits. There is an established cognitive-experimental literature exploring alternative methods for working with these distributions (see, for example, Balota & Yap, 2011). However, there are more fundamental reasons why in the particular case of writers producing spontaneous, multi-sentence texts, keystroke latencies are not normally distributed. This is a direct result of the fact that writing processes cascade, with the cognitive processes necessary to form the next word or words on the page often *but not always or completely* running in parallel with preceding output. Where this parallel processing occurs transitions across word and sentence boundaries are very rapid. However, there are occasions where the cascade of processes upstream from motor planning of finger movements is disrupted – for example then the writer struggles to find a spelling or loses the thread of their argument, the latencies are substantially longer. There are, therefore, at least two distinct data-generating processes underlying observed latencies, each resulting in a different distribution, one that captures rapid, fluent transitions, and one or more that captures the minority of cases (in competent writers) where the cascade is disrupted. When plotted together these give the appearance of a single distribution with strong positive skew.

One strategy for handling the complex distribution of latencies from spontaneous text production is to avoid reporting central tendency altogether and to just count latencies that exceed a specific threshold, traditionally set at 2 s. “Pause” counting, a practice inherited from early research in speech production (see Rochester, 1973 for a review), has been widely adopted by writing researchers, with studies dating back to at least the mid-1990s (Foulin, 1995, 1998). As an approach to understanding writing latencies this has two disadvantages. First, researchers have to make an a priori decision about where to set the pause threshold. Current understanding of the processes that underlie text production does not provide a strong theoretical basis on which to make this decision. So while a 2 s threshold undoubtedly captures an interesting distinction – processing that occurs in the range zero to 2 s is very likely to be qualitatively different from processing that takes more than 2 s – the same could be argued for any threshold a researcher might care to choose between perhaps 250 ms and 10 s (Chenu et al., 2014). Moreover, as we have already discussed, interpretation of latencies depends on location within the text. A pause threshold that captures activity associated with content and syntax planning at the start of a sentence is likely to miss activity associated with orthographic processing mid-word.

A second disadvantage of pause counting is that dichotomizing latencies discards much of the information that is captured within a keystroke log or digitized handwriting trace. We therefore require statistical methods that model the variance across the full range of keystroke latencies (or similar measures from handwriting). These are described in papers by Hall et al., and by Roeser et al. in this special issue (and see also Baaijen et al., 2012; Chenu et al., 2014; Li, 2021).

Finally, interpreting keystroke or pen movement latencies in spontaneous production is complicated by the fact that output is rarely entirely linear: The sequence of letters, words and sentences in the final text typically does not map directly onto the sequence in which they were produced by the writer. The extent to which this is true

varies depending on writer and task. However very few texts are produced without the writer engaging in some level of revision, even if just to correct typos or misspellings. For typewritten production this also means that writers are not always writing at the front edge of their text but instead jump around. For researchers aiming to infer cognitive process from keystrokes this means that latencies at a particular text location must be interpreted with reference to what happens next: The interkey interval before a sentence that is then written without editing has a different interpretation to the interval before a sentence that the writer modifies during production to change syntax or meaning. Given the cascading nature of the processing underlying fluent production, it is also necessary to interpret latencies in light of immediately preceding production. The interkey interval before a sentence that is written immediately after the preceding sentence has a different interpretation to the interkey interval before a sentence that is inserted following the writer jumping back within their text (see Hall et al., 2022).

A third reason why we might want to explore the writing timecourse data, quite apart from its importance in interpreting production latencies, is to capture and analyse the ways in which texts change and develop over the course of their production. Focus here is not on the order or duration of events, per se, but on how content and language develop over time. Both keystroke logging and versioning (capturing that state of the text at regular intervals during its composition; e.g., Lo Sardo et al., 2023) provide data that enable analysis of text development over time.

Researchers face two main issues here. First, these text-development data must be processed to provide meaningful summaries. Two common approaches to describe text development, include (non-)linearity and revision analyses, focusing on changes in the author's point of inscription (location of text production) over time or changes in the produced content, respectively. This (non-)linear production has been analyzed using graphs, such as the LS-graph (Eva Lindgren & Sullivan, 2002) or Inputlog's process graph (Leijten & Van Waes, 2013); as well as via global linearity measures, such as linear transitions between sentences (Baaijen et al., 2012). Revision analyses include counting the number of deleted characters and manual annotations of revision (e.g., Stevenson et al., 2006). However, most of these approaches require extensive manual annotation or inspection, and hence are not directly of practical use for large numbers of texts (but there are some exceptions, e.g., S-notation; Kollberg & Eklundh, 2002; or edit distances; Lo Sardo et al., 2023). This special issue provides two additional data-driven approaches to automatically identifying non-linearity and revision in text development (Buschenhenke et al., 2023; Conijn et al., 2021).

A second issue here is that text development analyses need to be able to handle unfinished compositions or snapshots of the text at different points in time, which may involve unfinished sentences or words, notes, and misspellings. For example, the notion of 'leading edge', or the outer boundary of text production, becomes complicated when there is a trailing whitespace or even some trailing text (e.g., a bibliography list), that the writer keeps pushing forward (Conijn et al., 2022a, b; Lindgren et al., 2019). This becomes even more complicated when writers are creating large texts, where they write different sections in a non-linear order (see Buschenhenke et al., 2023). Moreover, when the writer moves their cursor or starts to revise in the middle of an unfinished sentence or word, it becomes difficult or sometimes even

impossible to determine what the writer intended to write. This complicates analyses aimed at the content of the text, including (manual or automated) annotation and the use of natural language processing. For example, the type or orientation of revision may be hard to annotate when the writer started to revise in an unfinished word at the start of the sentence, making it unclear whether the writer intended a spelling revision, or larger, semantic change, changing the meaning of the text (Conijn, Speltz et al., 2022). Mahlow et al. (2022) provide a solution for parsing unfinished and ill-formed text.

Papers in this special issue

In what follows we briefly summarise and discuss each paper in this special issue, divided loosely into three groups: Papers that present tools for timecourse data capture or coding, papers that focus on the processing and statistical analysis of keystroke latencies, and papers that describe methods for studying non-linearity of written composition and how text develops over time.

Although the vast majority of adult writing, in most contexts, is by keyboard, most children learn to handwrite before they learn to type. G-STUDIO (Chesnet et al., 1994) was one of the earliest research tools for handwriting capture, replay and analysis. This grew into *Eye and Pen*, which is now in its third major version. The new functionality of this version is described in Chesnet et al. (2022). Readers not already familiar with *Eye and Pen* may want to start with Alamargot et al. (2006), or just explore the most recent version of the software (<https://eyeandpen.net/>) which is now available without payment. A main focus of this paper is a fairly technical description of timing issues associated with synchronizing timing between input devices – the digitizer on which participants write and an eye tracker if this is being used – and the computer used for data capture and, in the context of controlled experimental research, to display experimental stimuli. We believe that detail at this level (see also Hall et al., 2022) is important. There is a tendency for writing researchers to pass responsibility for the details of how they are capturing and processing timecourse data on to decisions made by the developers of the software they are using. Although to some extent this is inevitable, good science requires that researchers understand, own, and communicate important details of their tools and measures. In this regard the flexibility in choice of timing mode and fixation definition in the most recent version of *Eye and Pen* is very welcome.

Eye and Pen, as the name suggests, also supports synchronized collection of eye movement data. Eye movement has been exploited in a handful of writing timecourse studies to provide insight into the mental activity that occurs when writers pause during spontaneous production (Carl, 2012; Chukharev-Hudilainen et al., 2019; Torrance et al., 2016a). Collecting eye movement data from writers composing by keyboard, as is the case with in the three papers just cited, has the added advantage that with appropriate specialist software it is possible to automatically extract the text of the word or words that are fixated when a writer looks back into their text. This is exploited by the most recent version of *Scriptlog*, a keystroke logging program described in the contribution of Wengelin and co-workers to this special issue. Unlike

most previous studies that have sought to explain activity during pauses in typing, they illustrate the new Scriptlog functionality by describing keystroke activity that occurs during fixations.

Fitjar and co-workers provide the only other paper in this special issue that presents methods for studying handwritten production. They focus narrowly on interpreting digital traces generated during the production of isolated letters, a problem that must be solved by researchers interesting in development of graphomotor skills in children who are at the very start of learning to write. The problem that they aim to solve is an example of the unit-of-analysis problem that we discussed previously, but at the level of sub-letter graphical features rather than linguistic units that span one or more words. There is an established literature describing a range of measures of graphomotor fluency (see Danna et al., 2013 for a comprehensive summary). The challenge faced by researchers is to find units of analysis that permit comparison in pen-movement fluency. They argue that these necessarily must be defined just by required letter (allograph) features, independently of how these are produced, and provide an illustrative coding scheme and analysis that achieves this end.

There then follow three papers (Haake et al., 2022; Hall et al., 2022; Roeser et al., 2021) that describe statistical methods for interpreting keystroke data. Both Hall and co-workers and Roeser and co-workers specifically address the multiple-distributions problem that we detailed above, in both cases using a statistical technique called mixture modelling. Mixture models start from the assumption that each observed data point – each mid-word inter-keystroke interval, for example – belongs to one of two or more possible underlying distributions, each associated with a different data generating process. These distributions are then estimated. So, for example Hall et al. found that mid word latencies were best described by two distributions, one that captured the vast majority of keystrokes (around 95%) with a mean latency of around 137 ms, and another much smaller set of much longer latencies with a mean of just under half a second. In Roeser et al.'s terminology these represent fluent production – determined by just the time needed to move fingers to the next key – and hesitant production – what might traditionally be called “pausing” – where time between keys is determined by higher-level (pre-motor) processing (e.g., retrieving spelling). These two papers adopt quite different approaches to applying mixture modelling to their keystroke data. Hall et al. fit a mixture model separately to each participant whereas Roeser et al. adopt a linear mixed effects approach, with participant modelled as a random effect. Both approaches recognize variation in typing skill across writers – a criticism sometimes levelled at research that employs a fixed pause threshold. A combination of mixture and mixed-effects modelling, in particular, arguably provides a powerful and flexible tool for testing hypotheses about underlying cognitive processes both in experimental contexts and from studies of the production of spontaneous text.

Haake and co-workers (2022) offer an approach to interpreting keystroke data using a statistical technique called recurrence quantification analysis (RQA) to quantify regularity of keystroke intervals across a writing session. This approach identifies, at a participant level, whether the extent to which the writing timecourse exhibits temporal patterns in keystroke durations, described with several different summary statistics. For example, one measure of temporal patterning is to identify sequences

of interkey intervals that are of roughly the same duration and then find the average length of these sequences. Longer mean length indicates greater temporal regularity. Haake et al. (2022) demonstrate based on this and other RQA-derived measures, that there is greater temporal regularity in writers composing in their first language compared to writing in a language that they are learning. This finding is consistent with the findings from the mixture models reported in the two papers that we have just discussed. When writing is fluent, that is, when motor output is not delayed by upstream language processes, then keystroke intervals are short and regular. However, hesitation (pausing) can occur for a broad range of reasons, and therefore there is a much broader distribution of longer intervals.

The paper by Tian et al. (2021) goes beyond the analysis of keystroke data by describing how keystroke latencies may be linked to the writing product. This approach provides a valuable tool for instruction, where specific issues within the writing product may be linked to (suboptimal) strategies in composition. Correlating the writing product, including writing quality and overall cohesion, with the writing process is not new (see e.g., Conijn et al., 2022; Guo et al., 2018; Leijten et al., 2019; Sinharay et al., 2019). In this paper, Tian and co-workers go beyond this work by showing how natural language processing (NLP) can be used to extract more fine-grained linguistic features of cohesion, which in turn are related to a variety of writing fluency measures.

While Tian and co-workers use NLP on the final text, Mahlow et al. (2022) present an approach for applying NLP to the *evolving* written product. The added difficulty here, as described previously, is that the linguistic parsing must be done incrementally, with (usually) non-linearly produced text, and should be able to handle unfinished and ill-formed text. Mahlow and co-workers show how a syntactic parser, applied to raw keystroke data, may be used to explore the creation and revision of linguistic units, allowing for a better understanding of writing processes on a linguistic level. Moreover, their visualizations of text and sentence histories provide promising opportunities for real-time writing support. Both the papers by Tian et al. and Mahlow et al. demonstrate the added value of NLP in writing analysis for understanding composition at a linguistic level.

While Mahlow et al.'s approach can be considered as means of describing the development of the text at a sentence level, the approaches by Conijn et al. (2021) and Buschenhenke et al. (2023) are more activity-focused, focusing on the development of the text through, respectively, revisions and breaks in linear text production ('jumps'). Conijn and co-workers describe the automated extraction of what they term *revision events*, which include insertions and deletions at the leading edge, as well as deletions away from the leading edge. They show that machine learning can be used to automatically extract revision events from keystroke data without the need for manual annotation. With the use of NLP, the revision events in Conijn et al., or similarly the transforming sequences in Mahlow et al., could be further characterized, providing a replicable approach that can be applied across large amounts of text without the time, effort, and potential for error associated with manual coding.

Finally, Buschenhenke et al. apply a rule-based approach to detect and describe movements or jumps away from the point-of-utterance. This approach is applied to long-term multi-session writing, where non-linearity is a more complex construct to

define. In their proof-of-concept, the approach is applied to the composition of a full-length novel. The findings show how the characterization of jumps could be used to cluster writing sessions which are similar in terms of the non-linearity.

To conclude, this special issue describes a variety of methods for capturing and analysing writing timecourse data. At our request, as editors of this special issue, papers do not have as their main focus hypothesis testing or description of findings. They instead justify, describe in detail, and illustrate, specific new or underused approaches to understanding how and why text develops over time, providing where available open-source code and materials. We hope these methods will inspire and strengthen future empirical studies.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Ailhaud, E., & Chenu, F. (2017). Variations of chronometric measures of written production depending on clause packaging. *CogniTextes*, 17, <https://doi.org/10.4000/cognitextes.992>.
- Ailhaud, E., Chenu, F., & Jisa, H. (2016). A developmental perspective on the units of Written French. In J. Perera, M. Aparici, E. Rosado, & N. Salas (Eds.), *Written and spoken Language Development across the Lifespan: Essays in Honour of Liliana Tolchinsky* (pp. 287–305). Springer International Publishing. https://doi.org/10.1007/978-3-319-21136-7_17.
- Alamargot, D., Chesnet, D., Dansac, C., & Ros, C. (2006). Eye and Pen: A new device for studying reading during writing. *Behavior Research Methods Instruments & Computers*, 38(2), 287–299.
- 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>.
- Balota, D. A., & Yap, M. J. (2011). Moving beyond the mean in studies of mental chronometry: The power of response time distributional analyses. *Current Directions in Psychological Science*, 20(3), 160–166. <https://doi.org/10.1177/0963721411408885/FORMAT/EPUB>.
- Bonin, P., & Fayol, M. (2000). Writing words from pictures: What representations are activated, and when? *Memory & Cognition*, 28(4), 677–689.
- Bonin, P., Roux, S., Barry, C., & Canell, L. (2012). Evidence for a limited-cascading account of written word naming. *Journal of Experimental Psychology Learning Memory and Cognition*, 38(6), 1741–1758. <https://doi.org/10.1037/a0028471>.
- Braaksma, M. A. H., Rijlaarsdam, G., van den Bergh, H., & van Hout-Wolters, B. H. A. (2004). Observational learning and its effects on the orchestration of writing processes. *Cognition and Instruction*, 22(1), 1–36.
- 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.
- Buschenhenke, F., Conijn, R., & Van Waes, L. (2023). Measuring non-linearity of multi-session writing processes. *Reading and Writing*. <https://doi.org/10.1007/s11145-023-10449-9>.

- Carl, M. (2012). Translog-II: a Program for Recording User Activity Data for Empirical Reading and Writing Research. In N. Calzolari, K. Choukri, T. Declerck, M. U. Dogan, B. Maegaard, J. Mariani, J. Odijk, & S. Piperidis (Eds.), *LREC 2012 - EIGHTH INTERNATIONAL CONFERENCE ON LANGUAGE RESOURCES AND EVALUATION* (Issue 8th International Conference on Language Resources and Evaluation (LREC)), pp. 4108–4112.
- Chenu, F., Pellegrino, F., Jisa, H., & Fayol, M. (2014). Interword and intraword pause threshold in writing. *Frontiers in Psychology*, 5, <https://doi.org/10.3389/fpsyg.2014.00182>.
- Chesnet, D., Guillabert, F., & Espéret, É. (1994). G-STUDIO : Un logiciel pour l'étude en temps réel des paramètres temporels de la production écrite. *L'année Psychologique*, 94(2), 283–293. <https://doi.org/10.3406/psy.1994.28757>.
- Chesnet, D., Solier, C., Bordas, B., & Perret, C. (2022). A quick briefing on the new version of eye and pen (version 3.01): News and update. *Reading and Writing*. <https://doi.org/10.1007/s11145-022-10267-5>.
- Chukharev-Hudilainen, E., Saricaoglu, A., Torrance, M., & Feng, H. H. (2019). Combined deployable keystroke logging and eyetracking for investigating L2 writing fluency. *Studies in Second Language Acquisition*, 41(3), 583–604. <https://doi.org/10.1017/S027226311900007X>.
- Conijn, R., Cook, C., van Zaanen, M., & Van Waes, L. (2022a). Early prediction of writing quality using keystroke logging. *International Journal of Artificial Intelligence in Education*, 32(4), 835–866. <https://doi.org/10.1007/s40593-021-00268-w>.
- Conijn, R., Dux Speltz, E., & Chukharev-Hudilainen, E. (2021). Automated extraction of revision events from keystroke data. *Reading and Writing*, 1–26. <https://doi.org/10.1007/S11145-021-10222-W/TABLES/8>.
- Conijn, R., Roeser, J., & van Zaanen, M. (2019). Understanding the keystroke log: The effect of writing task on keystroke features. *Reading and Writing*, 32(9), 2353–2374. <https://doi.org/10.1007/s11145-019-09953-8>.
- Conijn, R., Speltz, E. D., van Zaanen, M., Waes, L., Van, & Chukharev-Hudilainen, E. (2022b). A product- and process-oriented tagset for revisions in writing. *Written Communication*, 39(1), 97–128. <https://doi.org/10.1177/07410883211052104>.
- Danna, J., Paz-Villagrán, V., & Velay, J. L. (2013). Signal-to-noise velocity peaks difference: A new method for evaluating the handwriting movement fluency in children with dysgraphia. *Research in Developmental Disabilities*, 34(12), 4375–4384. <https://doi.org/10.1016/j.ridd.2013.09.012>.
- Emig, J. (1971). *The composing processes of twelfth graders*. National Council for Teachers of English.
- Ericsson, K., & Simon, H. (1984). *Protocol analysis: Verbal reports as data*. MIT.
- Fitjar, C. L., Rønneberg, V., & Torrance, M. (2022). Assessing handwriting: A method for detailed analysis of letter-formation accuracy and fluency. *Reading and Writing*. <https://doi.org/10.1007/s11145-022-10308-z>.
- Flower, L., & Hayes, J. (1980). The dynamics of composing: Making plans and juggling constraints. In L. W. Gregg, & E. R. Steinberg (Eds.), *Cognitive processes in writing* (pp. 31–50). Erlbaum.
- Foulin, J. (1995). Pauses et débits : Les indicateurs temporels de la production écrite / pauses and rates: The temporal parameters of writing. *L'année Psychologique*, 95(3), 483–504. <https://doi.org/10.3406/psy.1995.28844>.
- Foulin, J. (1998). To what extent does pause location predict pause duration in adults' and children's writing? *Cahiers Do Psychologie Cognitive*, 17(3), 601–620.
- Gernsbacher, M. A., & Givón, T. (1995). *Coherence in spontaneous text*. John Benjamins Publishing Company.
- Graham, S., Harris, K. R., & Mason, L. (2005). Improving the writing performance, knowledge, and self-efficacy of struggling young writers: The effects of self-regulated strategy development. *Contemporary Educational Psychology*, 30(2), 207–241.
- Guo, H., Deane, P. D., van Rijn, P. W., Zhang, M., & Bennett, R. E. (2018). Modeling Basic writing processes from keystroke logs. *Journal of Educational Measurement*, 55(2), 194–216. <https://doi.org/10.1111/jedm.12172>.
- Haake, L., Wallot, S., Tschense, M., & Grabowski, J. (2022). Global temporal typing patterns in foreign language writing: Exploring language proficiency through recurrence quantification analysis (RQA). *Reading and Writing*. <https://doi.org/10.1007/s11145-022-10331-0>.
- Hall, S., Baaijen, V. M., & Galbraith, D. (2022). Constructing theoretically informed measures of pause duration in experimentally manipulated writing. *Reading and Writing*. <https://doi.org/10.1007/s11145-022-10284-4>.
- Hayes, J., & Flower, L. (1980a). Identifying the organisation of writing processes. In L. Gregg, & E. R. Steinberg (Eds.), *Cognitive processes in writing* (pp. 3–30). Erlbaum.

- Hayes, J., & Flower, L. (1980b). Writing as problem solving. *Visible Language, 14*(4), 288–299.
- Kollberg, P., & Eklundh, K. S. (2002). Studying Writers' Revising Patterns with S-Notation Analysis. In T. Olive & C. M. Levy (Eds.), *Contemporary tools and techniques for studying writing* (pp. 89–104). Springer. https://doi.org/10.1007/978-94-010-0468-8_5.
- Leijten, M., Van Horenbeek, E., & Van Waes, L. (2019). Analysing Keystroke Logging Data from a Linguistic Perspective. In Eva Lindgren, A. Westum, H. Outakoski, & K. P. H. Sullivan (Eds.), *Observing Writing. Insights from Keystroke Logging and Handwriting* (pp. 71–95). Brill. https://doi.org/10.1163/9789004392526_005.
- Leijten, M., & Van Waes, L. (2013). Keystroke logging in writing Research: Using Inputlog to analyze writing processes. *Written Communication, 30*(3), 358–392. <https://doi.org/10.1177/0741088313491692>.
- Levelt, W. J., & Maassen, B. (1981). Lexical search and order of mention in sentence production. In W. Klien, & W. Levelt (Eds.), *Crossing the boundaries in linguistics: Studies presented to Manfred Bierwisch* (pp. 221–252). Springer Netherlands.
- Lindgren, E., & Sullivan, K. P. H. (2002). The LS Graph: A methodology for visualizing writing revision. *Language Learning, 52*(3), 565–595. <https://doi.org/10.1111/1467-9922.00195>.
- Lindgren, E., Westum, A., Outakoski, H., & Sullivan, K. P. H. (2019). Revising at the Leading Edge: Shaping Ideas or Clearing up Noise. In Eva Lindgren, A. Westum, H. Outakoski, & K. P. H. Sullivan (Eds.), *Observing Writing. Insights from Keystroke Logging and Handwriting* (pp. 346–365). Brill. https://doi.org/10.1163/9789004392526_017.
- Li, T. (2021). Identifying Mixture Components from large-scale keystroke Log Data. *Frontiers in Psychology, 12*(July), 1–11. <https://doi.org/10.3389/fpsyg.2021.628660>.
- Lo Sardo, D. R., Gravino, P., Cuskley, C., & Loreto, V. (2023). Exploitation and exploration in text evolution. Quantifying planning and translation flows during writing. *PLOS ONE, 18*(3), e0283628. <https://doi.org/10.1371/journal.pone.0283628>.
- Mahlow, C., Ulasik, M. A., & Tuggener, D. (2022). Extraction of transforming sequences and sentence histories from writing process data: A first step towards linguistic modeling of writing. *Reading and Writing. https://doi.org/10.1007/s11145-021-10234-6*.
- Medimorec, S., & Risko, E. F. (2017). Pauses in written composition: On the importance of where writers pause. *Reading and Writing, 30*(6), 1267–1285. <https://doi.org/10.1007/s11145-017-9723-7>.
- Mohsen, M. A. (2021). Second Language Learners' pauses over different Times intervals in L2 writing essays: Evidence from a keystroke logging program. *Psycholinguistics, 30*(1), 180–202. <https://doi.org/10.31470/2309-1797-2021-30-1-180-202>.
- Newell, A. (1992). Precis of Unified theories of Cognition. *Behavioral and Brain Sciences, 15*, 425–492.
- Newell, A., & Simon, H. A. (1972). *Human problem solving*. Prentice Hall.
- Oldfield, R. C., & Wingfield, A. (1964). The time it takes to name an object. *Nature, 202*(4936), 1031–1032. <https://doi.org/10.1038/2021031a0>.
- Olive, T. (2014). Toward a parallel and cascading model of the writing system: A review of research on writing processes coordination. *Journal of Writing Research, 6*(2), 173–194. <https://doi.org/10.17239/jowr-2014.06.02.4>.
- Pinet, S., & Nozari, N. (2018). Twisting fingers: The case for interactivity in typed language production. *Psychonomic Bulletin and Review, 25*(4), 1449–1457. <https://doi.org/10.3758/s13423-018-1452-7>.
- Ronneberg, V., Torrance, M., Uppstad, P. H., & Johansson, C. (2022). The process-disruption hypothesis: How spelling and typing skill affects written composition process and product. *Psychological Research Psychologische Forschung, 86*(7), 2239–2255. <https://doi.org/10.1007/s00426-021-01625-z>.
- Rochester, S. R. (1973). The significance of pauses in spontaneous speech. *Journal of Psycholinguistic Research, 2*(1), 51–81. <https://doi.org/10.1007/BF01067111/METRICS>.
- Roeser, J., De Maeyer, S., Leijten, M., & Van Waes, L. (2021). Modelling typing disfluencies as finite mixture process. *Reading and Writing. https://doi.org/10.1007/s11145-021-10203-z*.
- Roeser, J., Torrance, M., & Baguley, T. (2019). Advance planning in written and spoken sentence production. *Journal of Experimental Psychology: Learning Memory and Cognition, 45*(11), 1983–2009. <https://doi.org/10.1037/xlm0000685>.
- Sinharay, S., Zhang, M., & Deane, P. (2019). Prediction of essay scores from writing process and product features using Data Mining methods. *Applied Measurement in Education, 32*(2), 116–137. <https://doi.org/10.1080/08957347.2019.1577245>.
- 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>.

- Tian, Y., Kim, M., Crossley, S., & Wan, Q. (2021). Cohesive devices as an indicator of L2 students' writing fluency. *Reading and Writing*. <https://doi.org/10.1007/s11145-021-10229-3>.
- Torrance, M. (2016). Understanding Planning in Text Production. In C. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of Writing Research* (2nd Editio, pp. 72–87). Guildford Press.
- Torrance, M., Johansson, R., Johansson, V., & Wengelin, Å. (2016a). Reading during the composition of multi-sentence texts: An eye-movement study. *Psychological Research Psychologische Forschung*, *80*(5), 729–743. <https://doi.org/10.1007/s00426-015-0683-8>.
- Torrance, M., Rønneberg, V., Johansson, C., & Uppstad, P. H. (2016b). Adolescent weak decoders writing in a shallow orthography: Process and product. *Scientific Studies of Reading*, *20*(5), 375–388. <https://doi.org/10.1080/10888438.2016.1205071>.
- Van Den Bergh, H., & Rijlaarsdam, G. (2001). Changes in cognitive activities during the writing process and relationships with text Quality. *Educational Psychology*, *21*(4), 373–385. <https://doi.org/10.1080/01443410120090777>.
- van Galen, G. P. (1991). Handwriting: Issues for a psychomotor theory. *Human Movement Science*, *10*(2–3), 165–191. [https://doi.org/10.1016/0167-9457\(91\)90003-G](https://doi.org/10.1016/0167-9457(91)90003-G).
- Wengelin, Å. (2002). *Text production in adults with reading and writing difficulties (Gothenburg Monographs of Linguistics, 20)*. Gothenburg University.
- Wengelin, Å., Johansson, R., Frid, J., & Johansson, V. (2023). Capturing writers' typing while visually attending the emerging text: A methodological approach. *Reading and Writing*. <https://doi.org/10.1007/s11145-022-10397-w>.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.